On Top of the Top: A Generalized Approach to the Estimation of Wealth Distributions

Franziska Disslbacher[a,b,c] Michael Ertl[d] Emanuel List[c]
Patrick Mokre[c,d,e] Matthias Schnetzer[d]

[a] University Roma Tre
[b] Stone Center on Socio-Economic Inequality, CUNY
[c] Research Institute Economics of Inequality, WU Vienna University of Economics and Business
[d] Department of Economics, Austrian Federal Chamber of Labour
[e] The New School For Social Research

Abstract

The wealth distribution is infamously top-heavy, while the decisive upper tail is missing from survey data on household wealth in European countries. We provide a novel quantile regression approach to estimate all parameters of the Pareto and Generalized Pareto distribution to adjust for the rich missing in survey data due to differential non-response and under-reporting. In contrast to existing Pareto-based adjustment routines, the generalized and rules-based method is scalable, flexible in the face of heterogeneities in data quality and wealth accumulation regimes, transparent, and prevents over-shooting of wealth aggregates and wealth concentration estimates. We apply the method to data on fourteen Eurozone countries by supplementing the Household Finance and Consumption Survey (HFCS) with a novel database on country-specific rich lists from the European Rich List Database (ERLDB) compiled from country-specific rich lists. The magnitude of the resulting upper-tail adjustments varies substantially across countries, highlighting the importance of the rules-based method developed here. In addition, while the results are highly stable across an extensive range of sensitivity tests addressing the opacities of ERLDB data, the resulting estimates vary substantially across parameters borrowed from prior work.

Corresponding author: Franziska Disslbacher, franziska.disslbacher@uniroma3.it. Address: Via Ostiense 161, 00154 Roma, Italy.

We thank Pablo Beramendi, Jürgen Essletzbichler, Salvatore Morelli, Mathias Moser, Emmanuel Saez, Martin Schürz, Thomas Plümper and Gabriel Zucman for helpful comments on this and an earlier version that has been circulated as *On Top of the Top - Adjusting Wealth Distributions Using National Rich Lists*. We also thank participants at the Inequality in Rome Seminar Series, the IIPF 2021 conference, the ECINEQ 2021 meeting, the IARIW conference 2021, and the Annual Congress of the Austrian Economic Association 2021 for their comments and feedback. Franziska Disslbacher thankfully acknowledges funding from the Austrian Economic Association, the Austrian Marshall Plan Foundation, and the Austrian Ministry of Education and Research. Patrick Mokre acknowledges financial support from the Austrian Chamber of Labor for the digitization of data.

Keywords: Pareto distribution, Generalized Pareto distribution, Wealth distribution, Inequality measurement, Distributional Financial Accounts, Top 1%
JEL Classification: C46, D31

1 Introduction

The size distribution of wealth is infamously top-heavy (Benhabib and Bisin, 2018) and the skewness of the wealth distribution results from a range of mechanisms (Jones, 2015; Piketty and Zucman, 2015; Benhabib et al., 2016; Gabaix et al., 2016). For the United States, the geographic focus of recent research on wealth inequality, different data sources and accompanying estimation methods have led to conflicting evidence on the level and trend of wealth concentration in the uppermost percentiles (Saez and Zucman, 2016; Saez and Zucman, 2020a; Smith et al., 2021; Smith et al., 2023). For most European countries, the very top of the wealth distribution is unobserved in what constitutes the most widely employed and only cross-country harmonized data source on household net wealth, namely wealth surveys (Vermeulen, 2016; Lustig, 2020; Ravallion, 2022). Due to the high concentration of wealth observed across countries and over time, it is precisely the upper tail that is decisive for the distribution of wealth across the total population (Piketty et al., 2022). The recent resurgence of interest in wealth taxation among economists and policymakers adds importance to the top tail (Seim, 2017; Saez and Zucman, 2019; Scheuer and Slemrod, 2020; Scheuer and Slemrod, 2021; Advani and Tarrant, 2021; Adam and Miller, 2021; Perret, 2021; Summers, 2021). Previous research has suggested Pareto-based methods to circumvent measurement problems resulting from the disproportionately low quality of survey data on the top of the wealth distribution (Vermeulen, 2016; Eckerstorfer et al., 2016). Such top corrections typically result in inequality measures that are substantially higher than those based on raw survey data. However, the previously proposed parametric and semi-parametric approaches rely on arbitrary specifications of crucial parameters, which is particularly problematic when combined with the assumption of uniform data quality across countries.

This paper introduces a novel and generalized quantile regression approach to the estimation of heavy-tailed distributions, specifically the Pareto and the more flexible Generalized Pareto (GP) distribution. Applying this methodology, we provide novel top-corrected and cross-country harmonized estimates of aggregate wealth and wealth inequality for 14 European countries. Our estimates rely on combined data from the Household Finance and Consumption Survey (HFCS) and a novel database on country-specific rich lists that we make publicly available as European Rich List Database (ERLDB, http://erldb.ineq.at). ERLDB constitutes the first cross-country collection of rich lists. In light of the absence of other data on wealth held by the super-rich, researchers increasingly use such lists to study the very top of the wealth distribution (Kaplan and Rauh, 2013; Alvaredo et al., 2018; Salach and Brzezinski, 2020; Luo and Chen, 2021; Advani et al., 2022; Moretti and Wilson, 2022; Tisch and Ischinsky, 2023; Baselgia and Martinez, 2023a). Typically, wealth levels reported in rich lists lie on top (of the top) of the survey distribution. Our generalized regression approach bridges the resulting lack of common support between the HFCS and ERLDB while taking neither source at face value and preventing over-shooting of rich-list-based estimates that prior work has documented (Kopczuk and Saez, 2004; Alvaredo et al., 2018; Baselgia and Martinez, 2023a). We find that the share of wealth held by the top 1% substantially increases in all countries when accounting for the underrepresentation of the upper tail. In the most extreme cases, it almost doubles. Overall, the magnitude of the tail adjustment varies substantially across countries and is closely related to cross-country variation in survey design and data quality. In addition, we find pronounced variation in the tail adjustment when we replicate previously employed strategies of arbitrary fixing crucial parameters. The variation across countries and the sensitivity to the parameter determination method highlight the importance of the transparent and rules-based regression approach we develop in this paper. By contrast, our approach results in highly stable outcomes across various sensitivity scenarios addressing the opacities of rich lists.

Our unified regression framework to estimate all parameters of the standard two-parameter Pareto distribution and the three-parameter Generalized Pareto distribution builds on Vilfredo Pareto's (1965) intuition that the upper tail of wealth distribution follows a power law. Pareto interpolation methods have been applied in the seminal contributions of Kuznets (1953), Atkinson and Harrison (1978), and Piketty and Saez (2003) and they are a crucial methodological ingredient of modern studies on top income and wealth shares (Atkinson and Piketty, 2007; Alvaredo et al., 2013; Föllmi and Martinez, 2017). We justify the Pareto distribution as a model for the upper tail of the wealth distribution since random growth models converge to a stable cross-sectional Pareto distributions. Early examples are Champernowne (1953) and Wold and Whittle (1957), while more recent micro-founded models account for both the level and trend of wealth inequality (Benhabib et al., 2011; Benhabib et al., 2015; Benhabib et al., 2016; Jones, 2015; Piketty and Zucman, 2015; Gabaix et al., 2016).[1] The standard Pareto distribution imposes strict linearity between the ranks of the wealth distribution and wealth levels. We address the concern that linearity might imply a too rigid wealth distribution model, especially in a cross-country setting, by incorporating and extending recent insights from Generalized Pareto (GP) modeling, thereby allowing for a drift-deviation from linearity (Atkinson, 2017; Blanchet et al., 2021; Jenkins, 2017; Blanchet et al., 2018).

Recent research on wealth concentration centers around five different types of microdata and corresponding methods and highlights that no single source can provide a consistent and comprehensive basis for the full support of the distribution. The first source is tax data resulting from the taxation of wealth or administrative wealth registers (Roine and Waldenström, 2015; Fagereng et al., 2016; Föllmi and Martinez, 2017; Jakobsen et al., 2020; Albers et al., 2022; Iacono and Palagi, 2023). Second, data on investment income streams are used in the capitalization approach (Saez and Zucman, 2016; Zucman, 2019; Saez and Zucman, 2020b; Saez and Zucman, 2020a; Smith et al., 2021; Smith et al., 2023; Garbinti et al., 2021; Saez and Zucman, 2022; Chatterjee et al., 2022; Martínez-Toledano, 2022). Third, data resulting from estate and inheritance taxation provide the basis for estimating the wealth of the living based on the wealth of the deceased (Kopczuk and Saez, 2004; Piketty et al., 2006; Roine and Waldenström, 2009; Alvaredo et al., 2018; Berman and Morelli, 2022; Acciari and Morelli, 2022). Fourth, surveys on household balance sheets have become an indispensable data source where administrative data is not available (Batty et al., 2021; Wildauer and Kapeller, 2022). Fifth, rich lists compiled and published by journalistic magazines provide insights into the super-rich's wealth (Klass et al., 2006; Bach et al., 2019; Brzezinski et al., 2020; Salach and Brzezinski, 2020; Tisch and Ischinsky, 2023; Baselgia and Martinez, 2023a). While research on the U.S. heavily draws on administrative data, most work on wealth inequality in Europe has relied on survey data. The few notable exceptions

[1] Examples for models on the distribution of income are Nirei (2009), Jones (2015), Nirei and Aoki (2016), Gabaix et al. (2016), and Jones and Kim (2018).

are Piketty et al. (2006), Föllmi and Martinez (2017), Alvaredo et al. (2018), Lundberg and Waldenström (2018), Jakobsen et al. (2020), Acciari and Morelli (2022), Garbinti et al. (2021), Albers et al. (2022), Martínez-Toledano (2022), and Iacono and Palagi (2023).

Estimating the level and trend in wealth inequality based on tax data involves several challenges. The stock of wealth is not taxed directly in most of the world (Kopczuk, 2015; Saez and Zucman, 2019; Scheuer and Slemrod, 2021). With the abandonment or suspension of wealth taxes during the last decades, administrative wealth tax data availability has deteriorated further (Saez and Zucman, 2020a; Scheuer and Slemrod, 2021). For a single country, estimates of wealth inequality can vary substantially across different types of tax data (tax on income streams, inheritance tax and wealth tax) and different assumptions imposed on one type of data. What and who is observed in tax data, in general, is determined by country-specific tax legislation, complicating comparisons of wealth inequalities across countries. For instance, there are decisive variations in the exemption threshold, the tax unit, the definition of the tax base and the reporting and valuation standards (Advani and Tarrant, 2021; Piketty et al., 2022). Consequently, estimating wealth inequality in multiple countries based on a similar concept of wealth is a crucial challenge far from resolved.

Wealth surveys fill critical data gaps but impose distinct challenges. On the positive side, they aim to capture the level and composition of net wealth of the total population based on (ex-ante) harmonized concepts and definitions. On the downside, wealth surveys are — as any survey — subject to sampling and non-sampling errors. The wealthiest households are less likely to be captured (correctly) than their lower percentile counterparts due to (1) a higher likelihood to refuse participation (Kennickell and Woodburn, 1999) and (2) more complex financial portfolios favoring misreporting, especially under-reporting (Vermeulen, 2016). Prior research has documented two types of wealth gaps resulting from survey errors for several countries. First, the micro-macro gap between aggregate wealth according to survey data and the assets recorded in macroeconomic balance sheets (Waltl and Chakraborty, 2022; Ahnert et al., 2020). Second, the micro-micro gap between the highest fortune according to a wealth survey and the smallest fortune according to a rich list (Eckerstorfer et al., 2016; Vermeulen, 2016; Wildauer and Kapeller, 2022). These gaps have been coined the problem of the "missing rich".

The growing literature on the extent of wealth concentration in European countries approximates the upper tail by using survey data as the lower and rich list observations as the upper bound to interpolate a Pareto distribution (Vermeulen, 2018). The goal of using the survey-rich list combination is to close both the macro-micro and the micro-micro gap while correcting the survey distribution for the missing rich.[2] Even though the methodologies underlying rich lists are opaque, the lists are still the best data source on wealth held at the very top (Piketty et al., 2022). In a seminal contribution introducing the combination of survey data and rich lists for eight European countries, Vermeulen (2016) finds that the share of wealth held by the top 1% is underestimated by between one (Spain) and eleven percentage points (Austria) in raw survey data. Subsequent research following the same approach provides quantitatively similar results for some countries and more pronounced tail adjustments for others (Eckerstorfer et al., 2016; Waltl and Chakraborty, 2022; Bach et al., 2019; Brzezinski et al., 2020).

In general, estimating the Pareto distribution hinges on two decisive parameters, and a third parameter allows for obtaining a top-corrected semi-parametric wealth distribution. The first parameter, w_{min}, locates the Pareto distribution. The second parameter, α, is the shape parameter of the distribution and describes inequality in the tail above w_{min}. Against the background of the micro-micro gap between survey data and rich list observations, a third parameter, w_0, helps to close the gap by obtaining a semi-parametric distribution spanning the entire range of wealth. The replacement threshold parameter w_0 determines a point in the wealth distribution above which survey observations become unreliable. Above this threshold,[3] survey data is deemed unreliable and replaced by data simulated based on α (Eckerstorfer et al., 2016).

While the literature proposes several estimators for the shape parameter α, it has thus far relied on best guesses or visual inspection of distributions to choose the location parameter w_{min} and the replacement threshold w_0. A typical choice of the location parameter has been

[2] One of the first contributions that fitted a Pareto distribution to the Forbes 400 list is Klass et al. (2006).

[3] In related literature that merges distributions across different sources, especially from survey data and tax data, a conceptually similar parameter is usually named as the merging point (Lustig, 2020; Blanchet et al., 2022).

one million in nominal national currency. As the location parameter is the starting point of the power law distribution and closely related to the shape of the Pareto distribution, such absolute values are particularly problematic when held fixed across countries that differ in terms of the shape of the wealth distribution. While one may allow for a variation in the location parameter across countries (or periods), the question of which rules or methods to rely on in the specification of context-specific w_{min} and w_0 is unresolved.

The core contribution of this paper is a flexible approach to estimating all required parameters, w_{min}, α, and w_0, without the need for any arbitrary decision. Our rules-based quantile regression approach is essential for coping with country-specific idiosyncrasies, particularly regarding the shape of the wealth distribution and data quality.

Our methodology improves and generalizes existing Pareto-based methods along multiple lines. First, in contrast to previous work, our generalized quantile regression approach does not require arbitrary choices on any parameter of the standard Pareto, the more flexible Generalized Pareto distribution, nor the replacement threshold parameter. Second, it is flexible and accommodates differences in data quality that arise from variations in the coverage of the upper tail or other idiosyncrasies. Third, previous work on top tail adjustments for income and wealth distributions has adopted either a replacement or a reweighting strategy (Hlasny and Verme, 2018; Flachaire et al., 2021; Ravallion, 2022). Our method uses a combination of both. A pure reweighting approach is insufficient for our purpose as to the extent the wealthiest are unobserved in survey data, reweighting cannot improve their coverage. Reweighting, however, ensures a stable total population before and after the top correction. Fourth, our methodology applies to estimating heavy-tailed distributions in general, including (capital) income, city size (Gabaix, 1999), and firm size (Luttmer, 2007). We thus add to the literature on the linearized estimation of power laws in economics (Gabaix, 2016). While the Pareto distribution and Pareto-based top corrections are essential for estimating top income and wealth shares, they are especially so in the context of Distributional National and Financial Accounts currently under development and implementation (Engel et al., 2022; Zwijnenburg, 2022; Alvaredo et al., 2020; Ahnert et al., 2020; Batty et al., 2021; Kennickell et al., 2021; Waltl, 2022). Finally, our approach is conservative because it puts complete trust in neither survey data nor rich lists. This property is desirable as Pareto estimations using rich lists have been shown to overshoot estimates based on tax data by far (Alvaredo et al., 2018).

We also contribute data on wealth inequality in two respects. First, we bring the first publicly available database on (country-specific) rich lists in the form of the European Rich List Database. Second, we provide comparable top corrected estimates of wealth aggregates and wealth inequality for 14 Eurozone countries for which both HFCS and ERLDB data are available, a much larger set of countries than in related work. These estimates are direly needed, since the HFCS constitutes the only primary data source for cross-country comparison of wealth inequality in Europe.

The paper is organized as follows: Section 2 is dedicated to a discussion of the two data sources. First, we discuss the HFCS and emphasize differences in the survey methodologies and top tail coverage across countries. Second, we present the novel European Rich List Database (ERLDB). Section 3 introduces our generalized regression approach to estimating heavy-tailed distributions. Section 4 presents our findings and highlights the new estimates of wealth concentration, aggregate wealth and aggregate wealth as compared to national accounts across 14 European countries. Section 5 evaluates the sensitivity of these findings, particularly regarding uncertainties behind the ERLDB data and the generalized regression approach. Section 6 concludes.

2 Data

Based on our generalized quantile regression framework to adjust for differential non-response and under-reporting, we provide comparable estimates of wealth inequality for 14 Eurozone countries. As the estimation of statistics on aggregate wealth and its distribution that are comparable across countries remains a crucial challenge, we leverage the major European survey on household finances, the Household Finance and Consumption Survey (HFCS). The HFCS includes ex-ante harmonized data on net wealth at the household level. To address the systematic bias in the HFCS in terms of top-tail coverage (Kopczuk, 2015; Waltl and Chakraborty, 2022; Lustig, 2020; Kennickell, 2021; Schulz and Milaković, 2021), we supplement it with our new collection of rich lists that we make available as European Rich List Database (ERLDB). We compiled the latter from several sources, constituting the first systematic database on country-specific rich lists.

Surveys on household finances struggle to effectively represent the upper tail of the wealth distribution (Vermeulen, 2016; Kennickell, 2019; Vermeulen, 2018; Lustig, 2020; Ahnert et al., 2020; Ravallion, 2022; Wildauer and Kapeller, 2022). The main reasons are coverage errors, differential non-response and differential under-reporting.[4] First, coverage errors result from a sampling frame that is not representative of the population. In contrast to its U.S. counterpart (the Survey of Consumer Finances), the HFCS is not subject to the sampling-based exclusion of the wealthiest. Second, prior work has documented non-response increasing with wealth and characteristics that are highly correlated with wealth (Davies and Shorrocks, 2000; Osier, 2016; Kennickell, 2019). Due to non-random non-response, estimates of aggregate wealth and wealth inequality based on raw survey data are biased.[5] Finally, there are strong concerns of differential under-reporting of net wealth, such that under-reporting of wealth increases with wealth (Vermeulen, 2018; Flachaire et al., 2021; Schulz and Milaković, 2021) and adds to the non-response bias.

We introduce the European Rich List Database (ERLDB) as a complementary data source to address the systematic errors behind the HFCS. The ERLDB provides estimates of wealth levels at the top of the wealth distribution for 23 European countries based on country-specific rich lists. The combination of HFCS and ERLDB constitutes the basis for applying our regression-based approach to estimating heavy-tailed distributions.

2.1 Household Finance and Consumption Survey (HFCS)

The HFCS provides detailed information on the level and composition of real and financial assets and liabilities at the household level (European Central Bank, 2020).[6] We employ its third wave, surveyed mainly in 2017. For most participating countries, the HFCS constitutes

[4] For an extensive review see Lustig (2020).

[5] As Ravallion (2022) illustrates, it is a widespread misunderstanding that the under-representation of the wealthiest automatically results in downward biased estimates of wealth inequality.

[6] The survey is coordinated by the European Central Bank (ECB) but conducted by national central banks.

the only micro-level data source on net wealth. While all countries conceptually survey the same assets and liabilities and aim at covering the total population, the survey methodologies differ substantially between countries, especially in terms of strategies implemented to improve the coverage of the top tail. We provide key information on the country-specific survey designs behind the HFCS and summary statistics in Appendix B, Table B.1.

As responding to the HFCS is not obligatory, unit non-response is a core concern.[7] Due to the strong correlation between unit non-response and wealth (D'Alessio and Faiella, 2002; Kennickell and Woodburn, 1999), effectively surveying households that belong to the top tail is an additional challenge. The reasons are manifold: Wealthy households are more likely to be absent for extended periods, they may live in several residences, and they are more likely and able to protect their privacy. Furthermore, perceived and actual time restrictions of wealthy respondents and reluctance to disclose information about their financial situation contribute to the disproportionately higher non-response rate at the top (European Central Bank, 2020). Against this backdrop, most central banks conducting the HFCS follow the established practice of oversampling households assumed to be wealthy in addition to stratified sampling (Kennickell, 2008; Bricker et al., 2016; Pfeffer et al., 2016). However, the quality and effectiveness of differential sampling efforts targeting the top tail vary considerably across the countries.

The overall success of oversampling to circumvent differential non-response depends on the ability to identify and interview households belonging to the top tail of the wealth distribution. In the HFCS, oversampling strategies resort to individual-level, household-level, or group-specific auxiliary variables correlated with wealth. In France, oversampling relies on individual-level wealth from administrative data. Other countries use administrative data on wealth-correlated concepts (income in Finland, the size of the primary residence in Portugal). Several countries conduct oversampling at the regional level. In Germany, households living in cities with high property prices and high-income municipalities obtain a higher sampling probability. In Belgium, the target of oversampling is households residing in regions with a

[7] Generally, the HFCS tries to tackle non-response by ex-ante adjusting the sampling probabilities across strata that differ according to their predicted response rates, resulting in adjustments of the household-specific survey weights.

high dispersion of personal income. In Ireland, regional oversampling is implemented based on a wealth index composed of home ownership rates and local property tax revenues. Three of the 14 countries in our sample do not even attempt to oversample the top tail. These are Austria, Italy, the Netherlands and Slovenia (European Central Bank, 2020).

Differential non-response can still outweigh oversampling, and the effective oversampling rate provides an intuition of the country-specific quality and success of oversampling. It measures the number of households in the (unweighted) sample with wealth above a certain percentile according to the weighted data. A sample with a relatively large number of affluent households and correspondingly small average weights indicates an effective oversampling strategy. Albeit the effective oversampling rate assumes that the weighted data provides an accurate representation of the wealth distribution, it is still a useful measure to characterize cross-country differences in the success of oversampling. Table B.1 in Appendix B presents the oversampling strategies and the effective top 5% oversampling rates by country. The effective oversampling rate of the top 5% ranges from -15% in Austria — with no oversampling — to +278% in France, where oversampling is based on administrative wealth tax data. This striking variation in the effectiveness of oversampling underscores that any comparison of wealth-related statistics based on raw HFCS data can imply misleading conclusions. Our generalized quantile regression approach is sensitive to such differences in data quality.

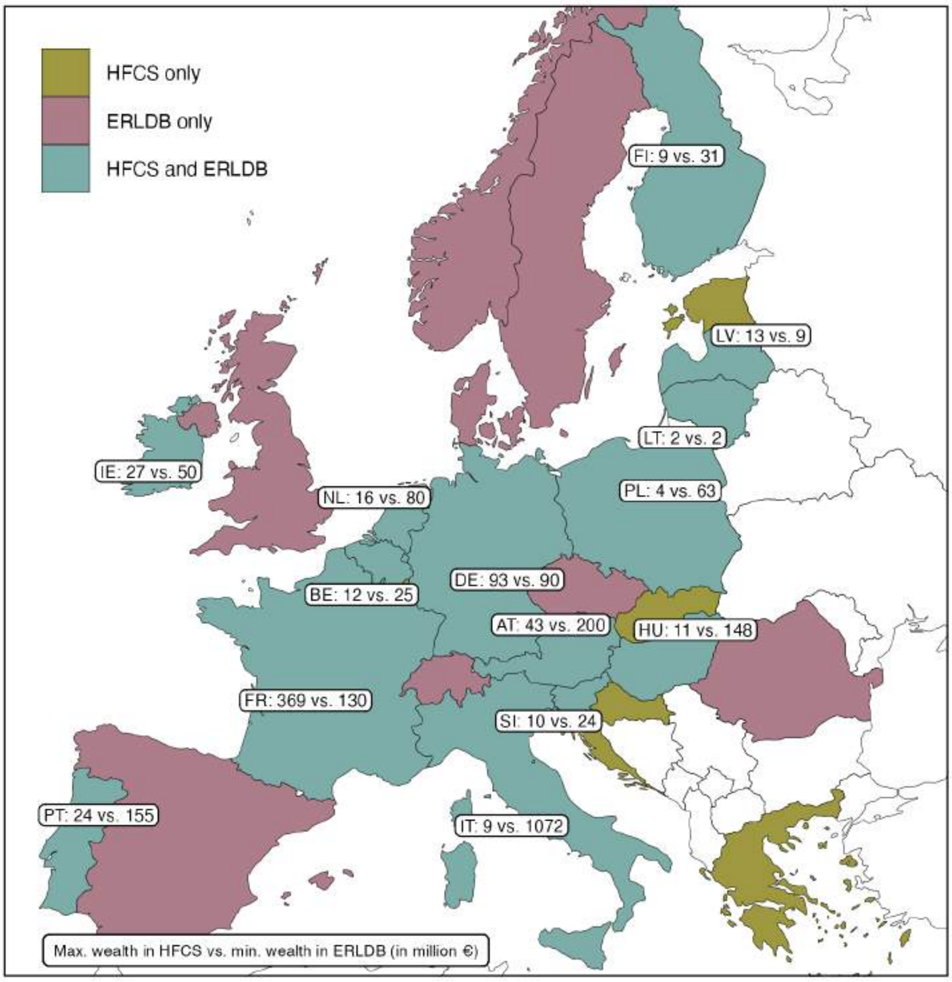
Responding households can refuse to answer single questions, for instance, if perceived as complex or sensitive, resulting in item non-response. In addition, a lack of information, recall problems, and a biased perception or memory of one's financial situation or the wish to conceal facts from an unknown interviewer may lead to factually wrong answers, i.e., under- or over-reporting. Both item non-response and misreporting are particularly problematic if they are not uniformly distributed along the wealth distribution, resulting in systematic biases. Regarding net wealth, a core concern is under-reporting which increases with wealth. For instance, wealth portfolios are increasingly complex towards the top, contributing to a disproportional prevalence and extent of under-reporting. Our methodology accounts for differential under-reporting by replacing wealth reported in the HFCS above the threshold w_0 with values derived from the parameter estimates of the (Generalized) Pareto distribution obtained from the combination of HFCS and ERLDB data.

All central banks responsible for the HFCS implement a multiple imputation strategy to keep item-non-responding observations in the HFCS sample. For each missing item, five estimates are provided, resulting in five implicates of the HFCS. We have calculated all estimates using Rubin's Rule (Little and Rubin, 2002); hence, they are the mean of estimates across five implicates.

2.2 European Rich List Database (ERLDB)

Despite the oversampling attempts behind the HFCS, aggregate household wealth according to the HFCS is, in most countries, considerably lower than corresponding aggregates from national accounts (Vermeulen, 2016; Waltl and Chakraborty, 2022). We supplement the HFCS data with country-specific rich lists that cover the wealthiest. Rich lists, on the one hand, are subject to methodological opacities which we discuss transparently. On the other hand, they provide essential information on individuals and families not captured in wealth surveys. To date, the lists are the best available data source on wealth at the very top (Piketty et al., 2022). In addition, the country-specific rich lists provide estimates of wealth levels for a much larger number of observations than the previously employed international lists, such as the Forbes list of billionaires.

We have collected rich lists for 23 countries, with roughly $13,300$ observations. We make the lists publicly available for research as the European Rich List Database (ERLDB, http://erldb.ineq.at). While the ERLDB is the first systematic collection of rich lists, researchers make increasingly use of such lists. For instance, Moretti and Wilson (2022) and Baselgia and Martinez (2023b) rely on rich lists to investigate behavioral responses the taxation among the wealthiest, Salach and Brzezinski (2020) to investigate the political connectedness of the superrich, Tisch and Ischinsky (2023) to understand the historical origins of top wealth, Baselgia and Martinez (2023a) to shed light on the socio-demography of the top tail of the wealth distribution in Switzerland and Advani et al. (2022) in the UK. Figure 1 shows the geographical coverage of the ERLDB and compares the maximum values in the HFCS with the minimum values in the rich lists. In Appendix B we provide information on the length of each rich list (Figure B.1) and the gap between HFCS and ERLDB by the length of the list (Figure B.2).



Note: This figure shows the geographical coverage of the European Rich List Database (ERLDB) and Household Finance and Consumption Survey (HFCS) 2017. The labels report the maximum wealth in the HFCS and the minimum wealth in the ERLDB in million €.

Figure 1: Survey-Rich List Gap and Geographic Coverage of the HFCS and ERLDB

While previous research using rich lists has almost exclusively worked with the international list of billionaires published by the U.S. magazine Forbes or the daily Bloomberg Billionaires Index, country-specific rich lists have some advantages. First, the Forbes list only contains U.S. Dollar billionaires, whereas country-specific rich lists compiled by national magazines or newspapers also comprise observations with less wealth. Second, country-specific rich lists provide a significantly larger number of observations, listing up to 1,000 observations for a single country. The Forbes list totals roughly 2,100 observations worldwide, and the Bloomberg Billionaires Index includes only 500 observations. Country-specific rich lists might thus improve wealth estimates based on Pareto models, particularly for countries with only a few (or even no) entries in the international lists (Bach et al., 2019). Addressing the impact of the length of rich lists on the accuracy of Pareto-based wealth inequality measures within a Monte-Carlo simulation, Wildauer and Kapeller (2022) show that long country-specific lists outperform the shorter international lists. Third, local journalists might have better insights, sources and intuition regarding the wealth portfolios of the ultra-wealthy in a specific country than an international team of journalists. Nevertheless, the country-specific lists are subject to methodological opacities that we address in large sets of sensitivity scenarios.

Rich lists suffer from opacities along three lines. First, it is questionable if rich lists are exhaustive. Specific individuals can opt out or do not appear for other reasons, even though they would qualify (Kennickell, 2003). Relatedly, concerns about opting in can be raised. The inclusion of individuals or households in a list might result from efforts to maximize the attention that the magazine publishing the list receives. Hence, some observations might be included, although they would not qualify given their wealth. Second, as journalists rely on publicly available information to compile a rich list, the estimated wealth levels may be flawed, particularly regarding assets and liabilities held outside of listed companies. More generally, the value of some asset classes is difficult to assess, for instance, valuables (e.g. art collections) and wealth held in non-traded corporations. Further, debt is less visible than assets, potentially causing net worth to be overestimated (Kopczuk, 2015; Atkinson, 2008; Davies and Shorrocks, 2000). Third, the unit of observation of a rich list is not homogenous along a list. In some cases, a single list reports wealth held by individuals, households, families, and multi-generational dynasties consisting of multiple households. In our baseline scenario, we assume that the unit of observation is the household. However, we address this and other limitations of the ERLDB, especially related to inclusion and exclusion criteria and reported wealth levels, in a large set of sensitivity scenarios that manipulate the lists accordingly.

Several papers have attempted to validate rich lists with a secondary source. Unfortunately, none of them refers to a country included in our sample. For the case of the U.S. Forbes 400 listing of the wealthiest Americans, Saez and Zucman (2016) have shown that net wealth according to the list is consistent with (confidential) IRS tax return data at the individual level. Likewise, Moretti and Wilson (2022) validated the Forbes 400 list based on estate tax revenues. By contrast, Kopczuk and Saez (2004) concludes that Forbes-based top wealth shares are substantially over-estimated compared to wealth shares derived from estate tax returns. Alvaredo et al. (2018) reach a similar conclusion for the case of the UK, pointing towards an over-shooting of estimates of wealth concentration if rich lists are taken at face value. In their comparison, both Kopczuk and Saez (2004) and Alvaredo et al. (2018) derive the (list-based) wealth share of the top 0.0001% using only rich lists data. By contrast, our generalized regression approach does not fully trust rich lists. It uses them as an auxiliary source to obtain a semi-parametric distribution located between the HFCS and ERLDB data.

For merging ERLDB and HFCS on a country-year basis, we have chosen the year of the rich lists closest to the HFCS reference period. In some cases, though, the interview period of the HFCS and the reference period of the rich lists do not overlap exactly, with a difference ranging up to several months. Additionally, the interview period of the HFCS was not restricted to a calendar year in some countries but spanned over two years. In these cases, we selected the rich lists corresponding to the HFCS reference year during which most of the HFCS interviews were conducted. Table B.1 presents detailed information on the number of observations and reference years of the ERLDB.

3 A Generalized Regression Approach to the Estimation of Heavy Tailed Distributions

While the HFCS suffers from differential biases, the country-specific rich lists in the ERLDB are subject to several methodological opacities. Our generalized regression approach tackles both problems. Overall, we propose a conservative approach that puts a share of trust in each of the two sources to prevent over-shooting of the resulting estimates of wealth concentration and wealth aggregates. In this section, we first introduce our generalized quantile regression approach to estimating the parameters of the (Generalized) Pareto distribution. Next, we define the transition threshold parameter we use to obtain a top-corrected wealth distribution.

We start by outlining the approach for the case of the two-parameter Pareto distribution, followed by the case of the more flexible Generalized Pareto distribution.

3.1 Pareto Distribution

Our method is based on the observation that the distribution of wealth takes a remarkably similar form across countries and periods, resembling a power law. It was 19th-century Italian economist Vilfredo Pareto (1965) who observed that the wealthiest 20% of the population owned 80% of Italian land and formulated *Pareto's Principle.* The subsequent generalization that wealth distributions follow a power law is controversial until today, particularly regarding the forces generating a heavy top tail. As our goal is to obtain a wealth distribution for the entire range of wealth, we treat the wealth distribution as a mixed distribution with a Pareto upper tail (Brzezinski, 2014; Clauset et al., 2009). In estimating the parametric top tail, correctly defining the lower bound of the Pareto distribution is key. A simple generalization of Pareto's power law distribution denotes

$$f(w \mid w_{min}, \alpha) = \frac{\alpha \; w_{min}^{\alpha}}{w^{\alpha+1}} \tag{1}$$

where w_i is the wealth of observation i and w_{min} is the lower bound of observations closely following the power law. The distribution obtains a linear relationship between the logarithm of the complementary cumulative distribution $log(1 - F(w_i))$ and the logarithm of wealth $log(w_i)$

$$1 - F(w_i \mid w_{min}, \alpha) = \left(\frac{w_{min}}{w_i}\right)^{\alpha} \tag{2}$$

$$log(1 - F(w_i \mid w_{min}, \alpha)) = \alpha log(w_{min}) - \alpha log(w_i) \tag{3}$$

Log-log plots of the CCDF against ranked observations reveal the characteristic linear pattern at a glance. The simplicity of detecting the presence or absence of linearity adds to the model's popularity. The recent empirical literature has thus rediscovered the Pareto distribution as an approximation for the top tail of wealth distribution (Davies and Shorrocks, 2000; Klass et al., 2006; Gabaix, 2016; Vermeulen, 2016; Campolieti, 2018; Bach et al., 2019). Moreover, the Pareto distribution is an essential ingredient of the seminal work by Kuznets (1953), Atkinson and Harrison (1978), and Piketty (2003) and of recent research that estimates long-run series of the distribution of income and wealth. Recently, the Pareto distribution also features prominently in the literature on Distributional National Accounts (Blanchet et al., 2021; Alvaredo et al., 2020).

We draw on these classical and recent studies on wealth concentration and extend them by presenting a unified estimation approach derived from the properties of the complementary cumulative density function (CCDF). Our approach avoids accumulating statistical uncertainties due to the combination of several methodologies, as has been the practice in past work. In addition, we combine the insights of reweighting and replacement approaches to top-correcting distributions (Hlasny and Verme, 2018; Lustig, 2020; Ravallion, 2022; Blanchet et al., 2022).

We exploit the linear relationship of the logarithms to apply linear regression but implement rank correction on the left-hand side (Gabaix and Ibragimov, 2011) to avoid bias towards the leading rank. While the workhorse estimator of the linearized Pareto equation is OLS, we use a median quantile regression approach (Koenker and Bassett, 1978), thereby adding robustness to outliers (Waltl and Chakraborty, 2022). We obtain robust point estimates for the shape parameter α conditional on location w_{min}. The regression equation is given by

$$\log((i-0.5)\frac{\bar{N_{fi}}}{\bar{N}}) = \underbrace{\log(\frac{\bar{N}}{N}) + \alpha \log(w_{min})}_{\text{constant}} - \alpha \log(w_i) \tag{4}$$

where i is a decreasing ranking with $i = 1$ indicating the richest household, N being the sum of total weights, $\bar{N}$ indicating the average weight ($\bar{N} = \frac{\sum_{j=1}^{n} N_j}{n}$ in the sample of size n), and $\bar{N_{fi}}$ denoting the average weight of the first i highest observations, the left-hand side hence is the rank-corrected CCDF. α gives the slope of the log-linearized plot and is the inequality parameter of the standard two-parameter Pareto distribution. A smaller α corresponds to higher inequality within the tail. Notably, α depends on w_{min}, a problem we tackle by exploiting the linear form of the regression equation.

3.1.1 Estimation of the Pareto Location Parameter w_{min} and the Pareto Shape Parameter α

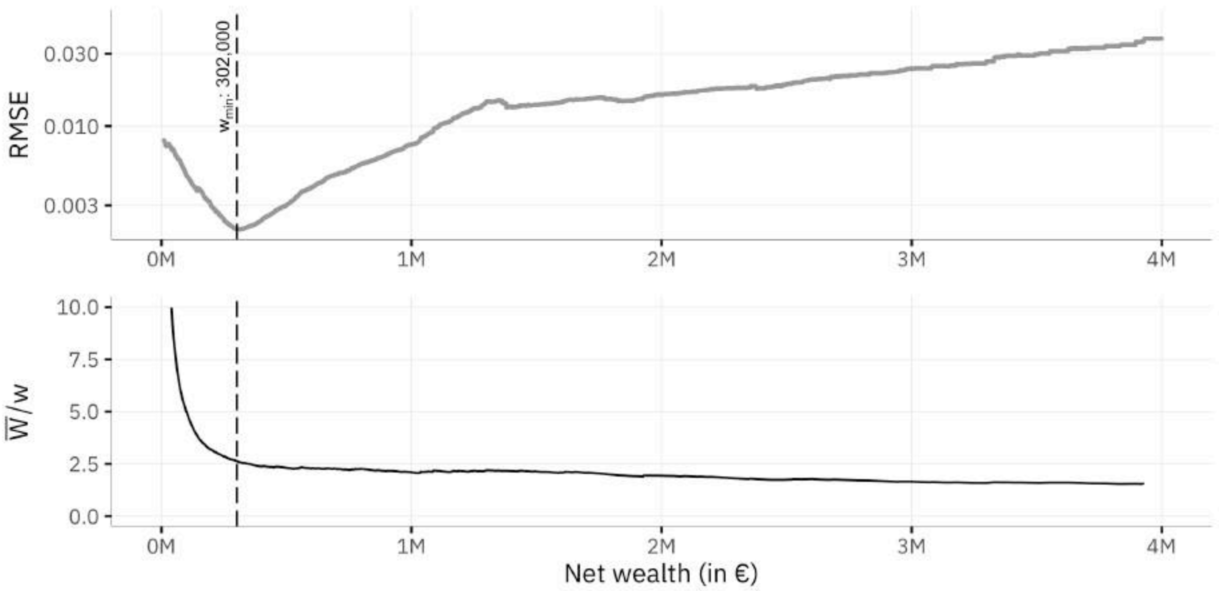
For each country, we estimate α in a median quantile regression corresponding to equation 4 based on all HFCS and ERLDB observations above location parameter w_{min}. As α depends on w_{min}, our choice of the location parameter w_{min} rests on the interpretation of the regression's root mean squared error (RMSE) as a measure of linearity. We thus algorithmically estimate w_{min} as the cut-off point above which observations follow the "most linear" CCDF-value relationship, i.e. we choose the RMSE-minimizing location parameter (Schulter, 2020).

The top panel of Figure 2 illustrates our process of estimating w_{min}. In steps of 1,000 €, we search for w_{min} between 0 € and 4 million € of net wealth. For each potential value of w_{min}, we estimate equation 4. Finally, we choose the w_{min} providing the minimum RMSE. The estimation of the regression equation for each potential w_{min} relies exclusively on HFCS data at and above w_{min}.[8] This restriction is motivated by the prevalent micro-micro gap between survey data and rich list observations. It ensures that our final estimates are located between HFCS and ERLDB data, and we return to this point at the end of this section. Given the RMSE-minimizing w_{min}, we re-estimate equation 4 based on HFCS and ERDLB data to obtain the final estimate of α. We provide figures on the RMSE-minimization process for all countries in Appendix C.

The bottom panel of Figure 2 reveals the distinctive property of the Pareto distribution known as *Van der Wijk's law*: the ratio of the average wealth of a subgroup above any threshold and the threshold itself is constant and determined by $\frac{\alpha}{1-\alpha}$, the inverted Pareto coefficient (Cowell, 2011). In previous work, the minimum of this ratio has served as a guideline for

[8] In addition, we require each regression to be based on at least ten observations. Our results show that increasing this minimum number of observations up to other meaningful limits will not impact the optimal choice of w_{min}.

choosing w_{min}. Comparing the bottom and top panels of Figure 2 lends further credibility to our regression-based estimation of the Pareto distribution. Another widely applied strategy circumvents choosing merely one w_{min}. Researchers frequently provide estimates of α for a small set of fixed location parameters (Vermeulen, 2016; Bach et al., 2019; Eckerstorfer et al., 2016), focusing on the covariation of w_{min} and α. Other studies suggest to choose the w_{min} that corresponds to the the minimum of the Kolmogorov-Smirnov distance metric between the empirical and theoretical distribution calculated for a set of candidate values of w_{min} (Clauset et al., 2009; Eckerstorfer et al., 2016). By contrast, we estimate a unique w_{min} and corresponding α, thereby our methodology does not rely on pre-specifying a small number of candidate values for the location parameter.



Note: The top panel of this figure shows our algorithmic estimation of the location parameter w_{min} based on the minimization of the RMSE. The search grid ranges from 0 to 4 million in steps of 1,000. We estimate the linearized Pareto equation for each value of w_{min} in the search interval. We choose the w_{min} providing the minimum RMSE and thus the most linear CCDF-value relationship in the HFCS data. Given w_{min}, we obtain α based on both HFCS and ERLDB data. The bottom panel illustrates *Van der Wijk's law* stating that the ratio between the average wealth above a given threshold and the threshold itself are constant if the data is Pareto distributed. The minimum of the ratio has been a popular choice of w_{min} in previous work. The figure is based on the first implicate of the HFCS 2017 for Germany.

Figure 2: Estimation of w_{min}

3.1.2 Transition Threshold w_0

The literature treats the gap between survey observations and rich lists as the result of differential under-reporting and non-response in the top percentiles of the survey data (Vermeulen, 2016; Vermeulen, 2018; Lustig, 2020). We introduce the parameter w_0, which indicates the point in the top tail above which the survey data is not trusted to be complete.[9] Our algorithm to determine w_0 rests on an argument advanced by Eckerstorfer et al. (2016) and Dalitz (2016): w_0 should coincide with the transition from continuous to discrete survey observations. We hence name it the transition threshold parameter. We locate w_0 as the point in the wealth distribution where the empirical density function of the data falls below the theoretical probability density function based on w_{min} and α. Equations 5 and 6 define the equality condition for w_0, which we determine (i.e., minimize) numerically.

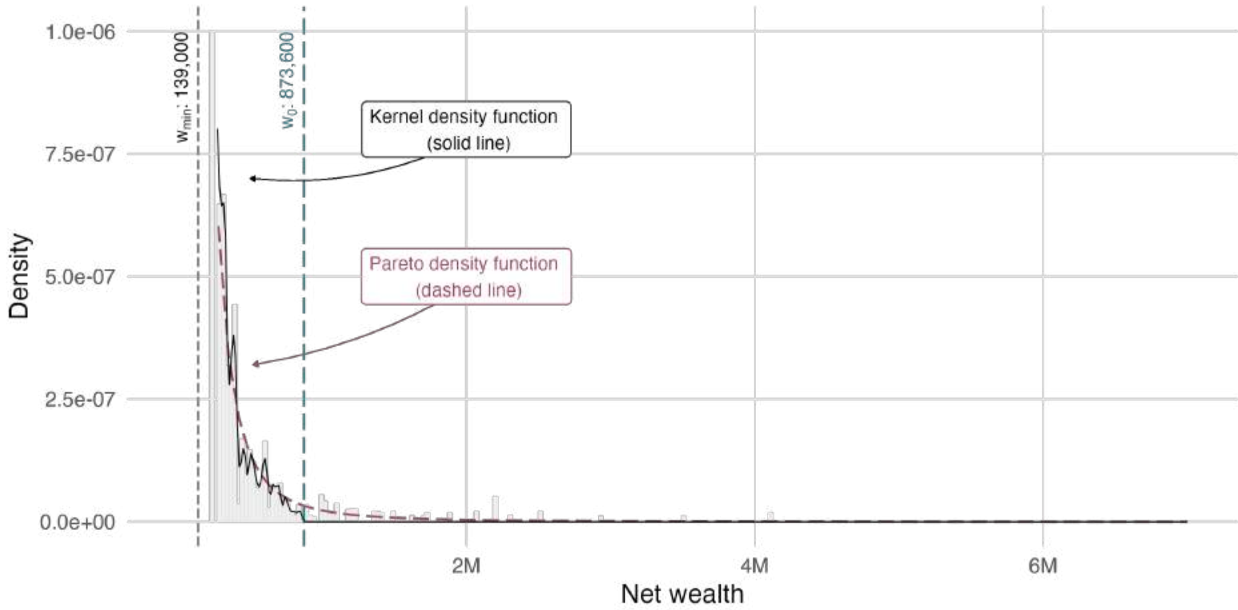
$$\hat{w}_0 = w_0 : \hat{f}_{kern}(w_0) = \frac{1}{Nh} \sum_i n(w_i) K(\frac{w_0 - w_i}{h}), \tag{5}$$

$$\hat{f}_{kern}(w_0) - \underbrace{\alpha\ w_{min}^{\alpha} \frac{1}{N} \sum_{w_i > w_{min}} n(w_i)}_{\text{normalizing constant C}} \times w_0^{-(\alpha+1)} = 0, \tag{6}$$

where $n(w_i)$ is the weight of household i, and h is the bandwidth for the kernel estimation, which we choose using the procedure proposed by Sheather and Jones (1991). Note that the equality condition for the theoretical and empirical density function includes a normalizing constant C. This constant adjusts the number of tail observations such that the sum of weights (the population size) before and after re-estimation remains the same (Eckerstorfer et al., 2016). C shifts the theoretical probability density function (PDF) up or down, which is crucial for finding the intersection of theoretical and empirical densities. Figure 3 illustrates the result for the case of Germany.

In practice, we locate w_0 as the point above which the empirical density function starts to continuously falls below the theoretical probability density. As the two density function may have multiple intersections, as illustrated in Figure 4, we proceed in four steps. First, we calculate the difference between the empirical and theoretical probability density function for each potential value of w_0. Here, we restrict the search to the interval $[\hat{w_{min}}, 10,000,000]$

[9] Approaches that focus on the reweighting of a survey-based distribution merged with a secondary source, particularly tax data, call a related parameter the merging point (Blanchet et al., 2022) since the weight of the data below (above) that point is decreased (increased) in the reweighting and merging process.



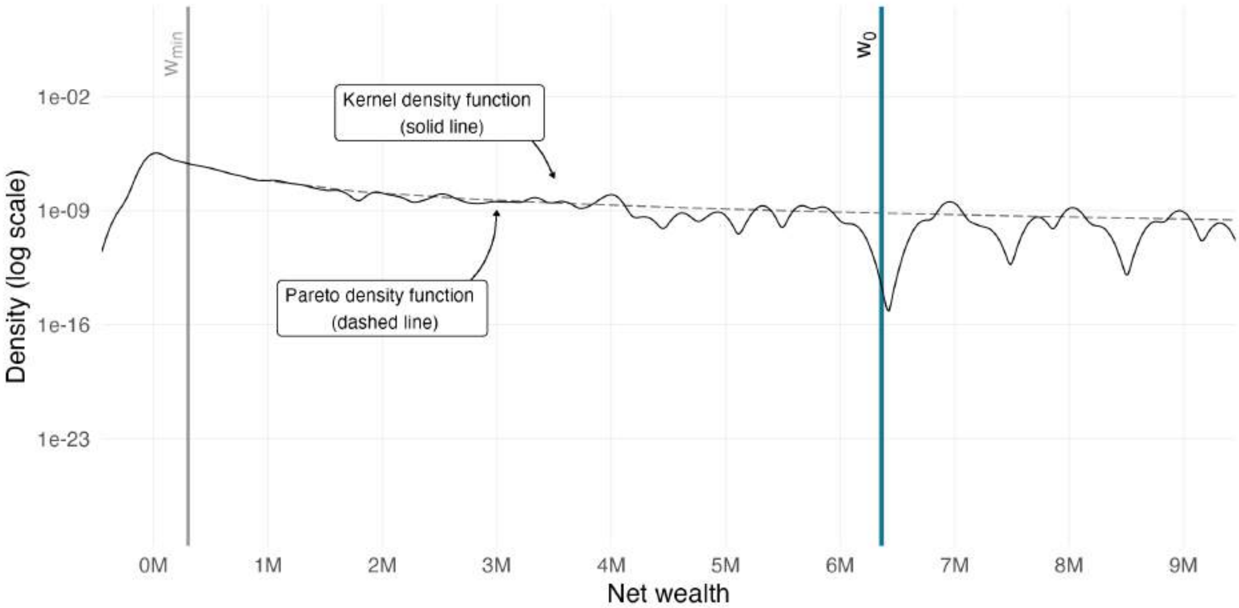
Note: This figure shows a histogram of the tail of wealth distribution above location parameter w_{min}, the kernel density function and the theoretical Pareto distribution. The transition threshold parameter w_0 provides the point in the wealth distribution above which survey data are no longer trusted to be complete. w_0 is the starting point for replacing survey data with observations drawn from the Pareto distribution. The figure is based on the first implicate of HFCS 2017 data for Germany.

Figure 3: Tail Histogram based on w_{min} and w_0

and we search for w_0 in steps of 100. Second, we compute the mean difference between the densities within 1,000 quantiles of the search interval. Third, we restrict the potential candidates for w_0 to the smallest 1% of the negative differences across the quantiles.[10] The purpose of steps two and three is to limit the influence of outliers on the difference between the densities resulting from the presence of a single survey observation. In addition, we thereby add the requirement of a continued (negative) difference over a certain interval. Finally, we pick the smallest possible value of w_0 among the remaining candidate values. This choice derives from searching for the point in the wealth distribution where the survey data starts to fall below the theoretical distribution. The last step is especially relevant in case of a constant difference along several quantiles around the search interval for w_0. We provide the figures illustrating the process of choosing w_0 for all countries in Appendix D.

We prefer this algorithmic approach to the visual inspection of functions since the latter is problematic for any cross-country, time-comparative, or multiple-implicate setting. Our

[10]The difference between the empirical and theoretical density function has to be negative in the range of possible values for w_0, we hence restrict the candidate values to the largest negative differences around a candidate value of w_0.



Note: This figure illustrates the algorithmic process of finding w_0, the transition threshold parameter. It shows the theoretical Pareto distribution density function and the kernel density function of the log of net wealth. It also illustrates the problem of multiple intersections of the two functions. We choose w_0 such that the kernel density function starts to fall continuously below the theoretical probability over a certain interval. For details, see the main text. The figure is based on the first implicate of HFCS 2017 data for Germany.

Figure 4: Determination of w_0.

unified estimation of w_{min} and α also reduces the sources of uncertainty. Furthermore, Dalitz (2016) points out that inequality measures of wealth distributions based on estimated Pareto tails vary substantially for different values of w_0. For this reason, we prioritize the transparent criterion suggested in this paper over the arbitrary determination as, for instance, in Eckerstorfer et al. (2016). Note that the distance between $\hat{w}_{min}$ and $\hat{w}_0$ is an indicator of how well surveyors were able to tackle differential biases among the wealthiest households. As the central banks participating in the HFCS employ substantially different oversampling strategies, we expect some variation in this distance, further emphasizing the need for our flexible and unambiguous procedure.

3.1.3 Pareto Tail

Finally, we obtain new observations above the transition parameter w_0 by simulation. We calculate the number of households with wealth above w_0 according to a $Pareto(\hat{\alpha}, \hat{w}_{min})$ distribution by extrapolating the number of households between w_{min} and w_0 with a cumulative density function above w_0 $(1 - F(\hat{w}_{min}))$. Thereby we obtain the theoretical share of tail observations above w_0. The tail length, which is the number of households above w_0, is defined by

$$\sum_{w_i > w_0} n(w_i) = [\sum_n (w_i)]_{w_i \in (w_{min}, w_0)} * \frac{1 - F(w_0)}{F(w_0)}. \tag{7}$$

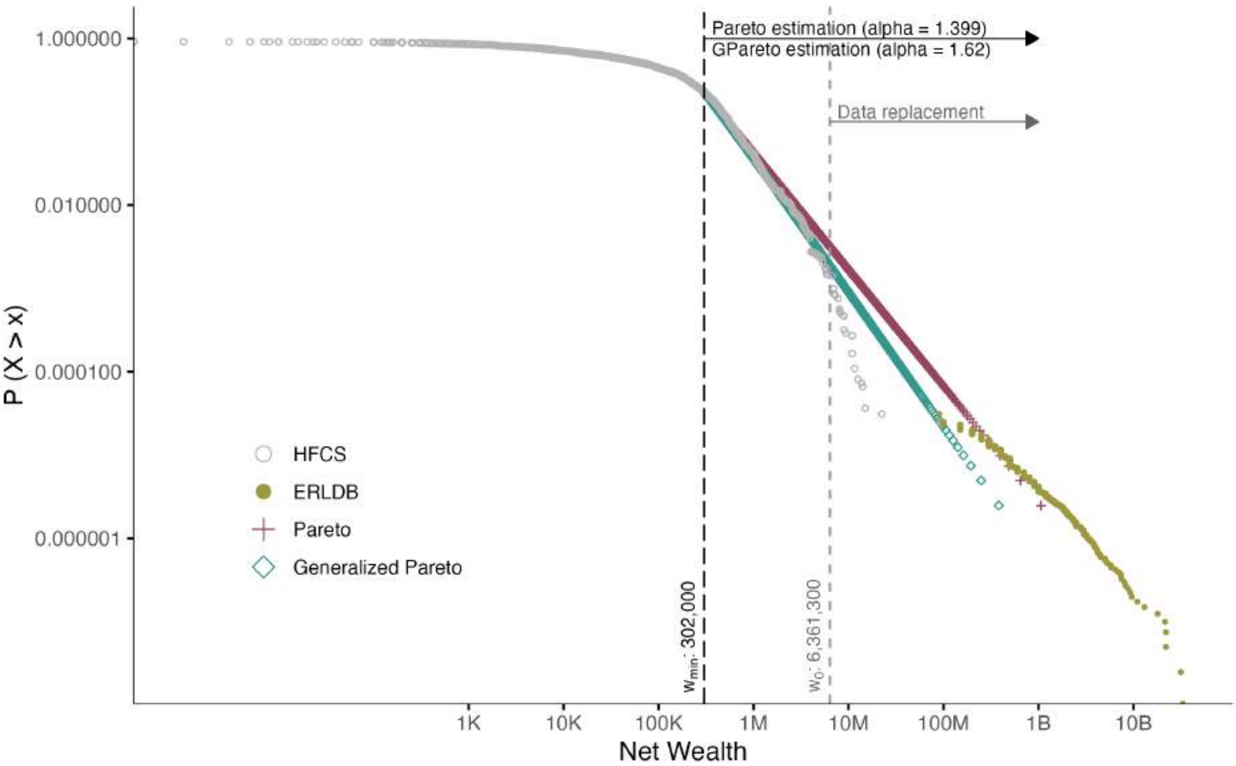
We rank the new simulated observations and assign net wealth according to

$$w_i = w_{min} \left(\frac{\sum_{w_i > w_{min}} n(w_i)}{\sum_{w_j > w_i} n(w_j)} \right)^{1/\alpha}. \tag{8}$$

Each of the simulated observations has a uniform household weight of 1. The combination of simulated observations and HFCS data below w_0 gives the re-estimated population. We linearly adjust the weights below w_0 to ensure that the re-estimated population corresponds to the target population in size. For the top-corrected distribution, we calculate inequality metrics, such as the share of wealth held by the top 1%, top 5%, and top 10%, P99/P50 quantile ratio, and the Gini coefficient.

We summarize the individual steps of our methodology toward a top-corrected wealth distribution in Figure 5. Based on HFCS data, we find the location parameter w_{min} that minimizes the RMSE of the linearized Pareto equation. The location of the distribution is hence chosen to result in the most linear CCDF-value relationship. Given w_{min}, we estimate α based on both HFCS and ERLDB data. We obtain w_0 as the point where the empirical probability density function starts to fall below the theoretical density continuously based on $\hat{w_{min}}$ and $\hat{\alpha}$. We finally obtain a top-corrected wealth distribution by ensuring that the population size remains constant. We treat the resulting distribution with survey observations up to $\hat{w}_0$ and simulated values above $\hat{w}_0$ as a distribution that corrects for differential biases.

Figure 5 depicts how our approach is conservative and prevents over-shooting of the estimate of the shape parameter α and the resulting adjustments in aggregate wealth and wealth inequality. First, our algorithm for detecting w_{min} rests only on HFCS data. If this process was based on both HFCS and ERLDB data, it would result in a higher estimate of w_{min} and, consequently, likely a much lower estimate of α. Second, the final quantile regression for estimating α uses both sources. For this reason, our approach always ends up fitting a distribution located between ERLDB and HFCS data. Even in the absence of the survey-rich list gap, our method puts a share of trust in both HFCS and ERLDB data, and the resulting parameter estimates will result in a Pareto distribution located between the two sources. Finally, the rich list is replaced by observations based on the estimated Pareto upper tail.



Note: This figure shows the complementary cumulative density function for the HFCS 2017 and the ERLDB data for Germany and the resulting estimates of the (Generalized) Pareto distribution. Based on the parameter estimates, we simulate new wealth observations above location parameter w_{min}. Survey observations above the transition threshold w_0 are replaced by wealth levels derived from the parametric distribution.

Figure 5: Complementary Cumulative Density Function of HFCS, Rich List, and (Generalized) Pareto Estimation and Simulation

3.2 Generalized Pareto Approach

Pareto's law approximates the tail of observable phenomena surprisingly well, but the simplicity of the two-parameter distribution implies rigidity. Atkinson (2017) stressed that Vilfredo

Pareto envisioned a richer functional form for the upper tail that requires rejecting a constant shape parameter α. In this spirit, Blanchet et al. (2018) and Blanchet et al. (2021) use a non-parametric definition of power laws to implement Generalized Pareto curves with varying α values along the distribution to interpolate tabulations of exhaustive tax data with a Generalized Pareto top tail for the uppermost bracket. By contrast, we rely on survey data but improve the functional form of the standard Pareto distribution by estimating a Generalized Pareto (GP) distribution for the top. The GP distribution is more flexible as it is defined by a three-parameter complementary cumulative density function (CCDF) as in

$$1 - F(w \mid, \xi, \mu, \sigma) = \left(1 + \xi \frac{w - \mu}{\sigma}\right)^{\frac{-1}{\xi}} \tag{9}$$

with a location parameter μ, shape parameter ξ, and scale parameter σ for $1+\xi(w-\mu)/\sigma > 0$ and $w > \mu$, where $\sigma > 0$. The shape parameter ξ relates to Pareto's α such that $\xi = \frac{1}{\alpha}$ (Jenkins, 2017). The location parameter μ has the same interpretation as w_{min}. As in the simple Pareto case, w_{min} indicates the threshold above which wealth approximately follows a GP distribution. We adopt the standard Pareto notation and use α_{GP} and w_{min} rather than ξ and μ since the two parameters share their interpretation. The scale parameter σ determines the drift towards the end of the tail and defines a higher or lower wealth concentration compared to the two-parameter Pareto distribution, which is a nested case of the GP distribution with $w_{min} = \frac{\sigma}{\xi}$ and therefore no drift from linearity by definition.

Our GP approach is an extension of efforts to approximate the top tail of wealth distribution. We build on the already detected threshold from the standard Pareto approach because the parameter shares its interpretation across the two distributions. We estimate the scale and shape parameters for a given w_{min}. Our estimation of the GP distribution's parameters builds on the insight that, if the scaled excesses of a random variable over a location parameter w_{min} follow a GP distribution, the scaled excesses for any threshold $u \geq w_{min}$ are also GP distributed with the same shape parameter $\frac{1}{\alpha_{GP}}$ (Langousis et al., 2016). Furthermore, the scale parameter σ_u depends linearly on the scale parameter of the threshold w_{min}, the shape parameter, and the excess over u. The scaled excess of a random variable over any threshold u is defined as $e(u) = E[W - u \mid W > u]$. Equation 10 gives the linear relationship for σ_u, equation 12 the expected value of the excess over u.

$$\sigma_u = \sigma_\mu + \frac{1}{\alpha_{GP}}(u - w_{min}) \tag{10}$$

$$e(u) = E[W - u \mid W > u] \tag{11}$$

$$= \frac{\sigma_u}{1 - \frac{1}{\alpha_{GP}}} \tag{12}$$

$$= \frac{\sigma_\mu + \frac{1}{\alpha_{GP}}(u - w_{min})}{1 - \frac{1}{\alpha_{GP}}} \tag{13}$$

$$= \beta_0 + \beta_1 u \tag{14}$$

The linear relationship in equation 12 allows for a linear regression estimation of both the scale and shape parameters, since $\beta_1 = \frac{1}{\alpha_{GP}}/(1 - \frac{1}{\alpha_{GP}})$ and $\beta_0 = (\sigma_u - \frac{1}{\alpha_{GP}} w_{min})/(1 - \frac{1}{\alpha_{GP}})$. Then, $\frac{1}{\alpha_{GP}} = \beta_1/(1 + \beta_1)$ and $\sigma_{w_{min}} = \beta_0(1 - \frac{1}{\alpha_{GP}}) + \frac{1}{\alpha_{GP}} w_{min}$.

We estimate the weighted mean excesses $e(w) = E[W - u \mid W > u]$ above different thresholds $u_i = W_{i,n}$ with $i = 1, 2, ..., n - 20$. Omitting the last (i.e., largest) 20 observations ensures that mean excesses are calculated based on at least 20 observations. This effectively pairs every observation w_i with a mean excess value $e(w_i) = E[W - w_i \mid W > w_i]$. For each observation w_i, $i = 1, 2, ..., n - 20$, we calculate the conditional weighted excess variance $Var[W - w_i \mid W > w_i]$ to account for the increasing estimation variance of $e(w_i)$ in w_i. We calculate the weights as $v_i = (N - i)/(Var[W - w_i \mid W > w_i]$. Finally, we perform a median quantile regression corresponding to equation 12, using v_i as weights.

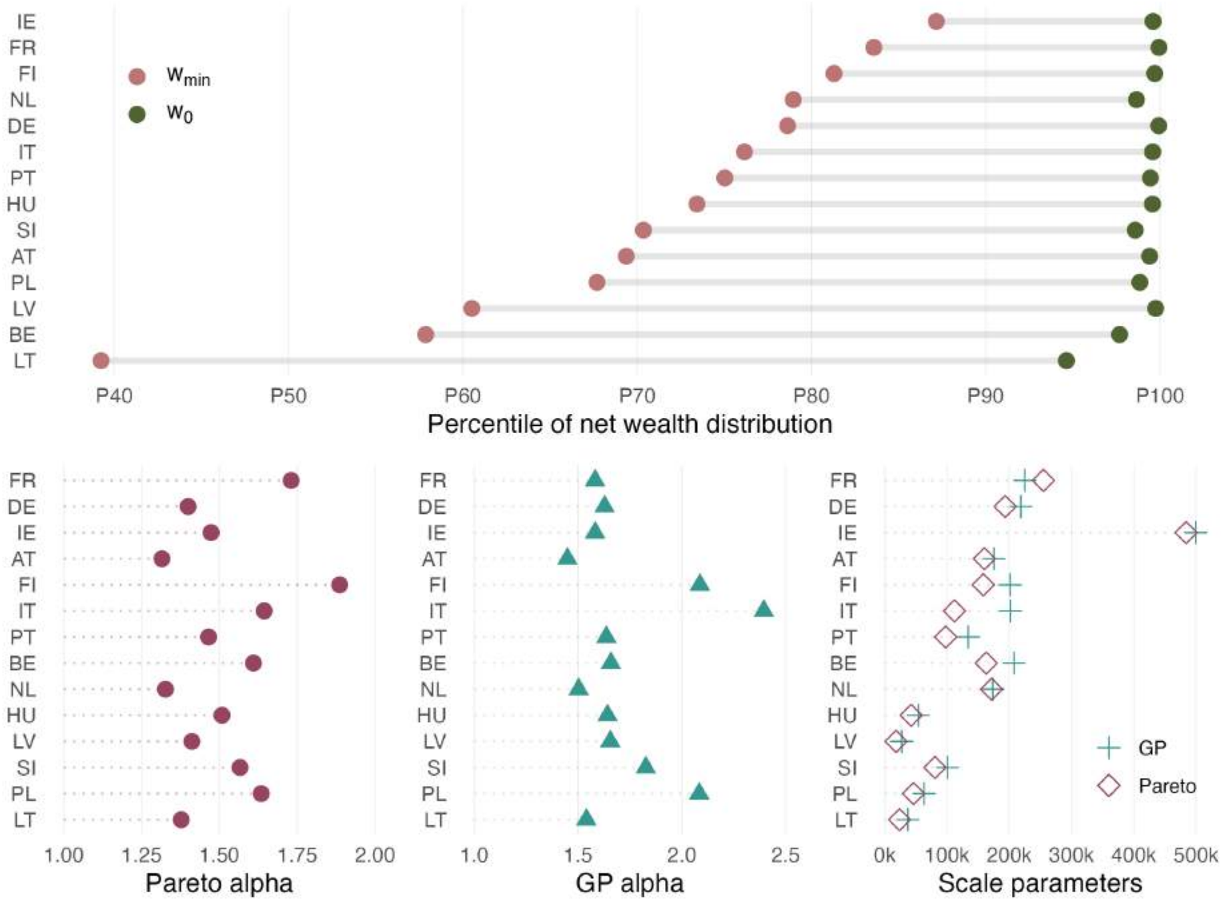
Our method detects the transition parameter w_0 where the empirical density suggests that we should no longer trust the survey data. Therefore, we use the same estimate of w_0 as in the case of the Pareto approach. Also, the calculation of the tail length of the GP distribution follows the same logic outlined for the case of the Pareto distribution. Again, as a last step, we simulate the tail above w_0 according to our estimates and assign wealth values to the simulated observations as in $GPareto\,(\hat{\alpha}_{GP}, \hat{\sigma}, \hat{w}_{min})$.

$$w_i = w_{min} + \alpha_{GP}\sigma \left[\left(\frac{\sum_{w_i > w_{min}} n(w_i)}{\sum_{w_j > w_i} n(w_j)} \right)^{-1/\alpha_{GP}} - 1 \right]. \tag{15}$$

We compare the results between the two different models for the top tail and demonstrate whether the drift deviation of the more flexible Generalized Pareto distribution outweighs the simplicity of the standard Pareto approach.

4 Results

We tackle both differential non-response and differential under-reporting in the combination of HFCS and ERLDB data and estimate cross-country comparable measures of aggregate wealth and wealth inequality for 14 European countries. We introduce a generalized and unified quantile regression approach to the (Generalized) Pareto distribution and incorporate recent findings from the literature on linearized parameter estimation of heavy-tailed distributions. While the Pareto approach allows us to close the gap between survey and rich list observations, we also estimate a three-parameter Generalized Pareto distribution. The latter is able to capture a drift deviation from the linear relationship between the logarithms of the complementary cumulative distribution function (CCDF) and wealth levels that is characteristic for the Pareto distribution. The GP approach entails a trade-off. While the distribution is more flexible and robust, especially when differential under-reporting is prevalent, it is more complex, and parameter estimation is more arduous than in the case of the simpler Pareto distribution. We obtain a location parameter w_{min} marking the threshold above which the data follows a Pareto distribution, and a shape parameter α that captures the degree of inequality in the tail. First, we apply median quantile regressions with a rank correction to determine point estimates of α over a sequence of w_{min}s. Then, we minimize the regressions' root mean squared error $RMSE(w, \alpha \mid w_{min})$ to obtain the corresponding parameters. Finally, we obtain a top-corrected wealth distribution by estimating the transition threshold parameter w_0. In the remainder of this section, we first discuss the parameter estimates of the Pareto distribution, followed by a comparative presentation of the results based on the GP distribution as a model for the top tail.



Note: This figure presents the parameter estimates of the Pareto and Generalized Pareto distributions and the transition threshold parameter w_0. The top panel shows the estimates of the location parameter w_{min} and transition parameter w_0 in terms of the corresponding percentile of the wealth distribution. The three bottom panels show the estimates of the shape and scale parameters.

Figure 6: Parameter Estimates (Generalized) Pareto distribution

4.1 Parameter Estimates of the Pareto and Generalized Pareto Distributions

We find considerable variation in the estimated location parameter w_{min} across countries. We locate the starting point of the Pareto distribution between the bottom 40% and the top 15% of the net wealth distribution, as illustrated in Figure 6. A full list of the corresponding estimates is provided in Appendix B, Table B.2. For Lithuania, the starting point of the Pareto tail is as low as the 39th percentile (€36,400) of the national net wealth distribution. We locate the Pareto distribution in Ireland at the 87th percentile (€765,600). For most other countries, our estimate of w_{min} is located between the 70th and 85th percentile of the wealth distribution, corresponding to substantially different absolute values. The wide range of location parameters indicates a considerable variety of wealth accumulation regimes in Europe and mirrors different oversampling strategies. The variety of best-fit location parameters also underlines the advantage of a unified and rule-based approach over arbitrary choices of w_{min}, especially when dealing with a cross-country data set.

We also find substantial variation in the shape parameter α, reflecting differences in the extent of wealth inequality. The lower α, as presented in the bottom left-hand panel of Figure 6, the higher inequality within the tail and, for a given location parameter, the higher inequality across the total population. The estimates of the shape parameter range from 1.32 in Austria to 1.89 in Finland. This finding is consistent with the assertion in Gabaix (2016) that parameter values around 1.5 are the norm for wealth distributions. The estimates of the α are also in the range of values presented in related work (Kapeller et al., 2021; Vermeulen, 2018; Brzezinski et al., 2020), even though our sample is a different HFCS wave and despite the application of a different estimation strategy.

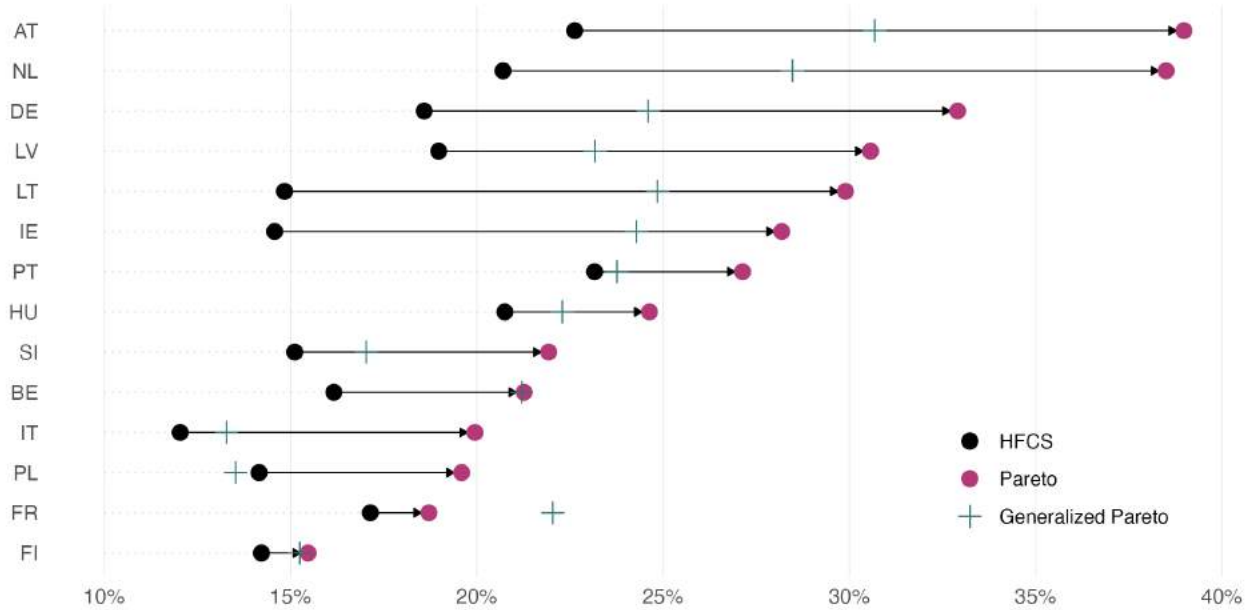
To obtain a top-corrected wealth distribution, we rely on the transition threshold w_0. Above this threshold, we disregard the empirical data and simulate observations based on the estimates of α and w_{min}. The position of the transition threshold w_0 in the net wealth distribution reflects the success of oversampling strategies to tackle differential non-response in the survey data and the quality of survey data more generally. The better the coverage of the top tail in survey data, the higher in the distribution we locate $\hat{w}_0$. We find a substantial correlation between the estimated $\hat{w}_0$ and the effective HFCS oversampling rate of the top 5% as shown in Figure B.3, Appendix B. Successful oversampling strategies imply a significantly lower fraction of simulated top-tail observations.

For the GP distribution, we rely on the shared interpretations of w_{min} and w_0 across the two parametric models. We use the same estimates of w_{min} and w_0 for the Pareto and GP distribution. w_{min} is the threshold where the data start to follow a (Generalized) Pareto distribution. w_0 is the point in the net wealth distribution beyond which differential non-response and under-reporting render the survey data implausible. The flexibility of the GP distribution stems from the scale parameter σ that determines the drift in the tail. When $\sigma = w_{min}/\alpha_{GP}$, the GP distribution equals a Pareto distribution. For a given α_{GP}, a scale parameter $\sigma < w_{min}/\alpha_{GP}$ implies that the heaviness of the tail increases towards the top, resulting in an increasing inequality along the tail. We present the estimates of the shape and scale parameters of the GP distribution in the bottom right-hand panels of Figure 6. In most countries, the scale parameter is close to the Pareto equivalent but somewhat higher. As a result, the heaviness slightly decreases towards the top of the tail in the GP framework. The simple Pareto distribution cannot pick up to such variation in inequality along the tail. Only in the case of France, the scale parameter of the GP distribution is lower than that of the Pareto distribution. The heavier GP tail in France is coherent with figure C.5, showing that the survey and rich lists data tend to form a convex curve on the CCDF plot.

4.2 Wealth Inequality

We sample wealth observations above the transition parameter w_0 based on the parameter estimates of the (Generalized) Pareto distribution by proceeding in two steps. First, we calculate the fraction (and number) of the population that belongs to the tail above w_0 using the cumulative density function. Next, we assign the appropriate theoretical quantile to each tail observation. We combine the simulated tail with survey observations and derive inequality measures and top wealth shares for the Pareto and Generalized Pareto distribution. In our primary analysis, we use this combination of HFCS observations and sampled data to obtain measures of wealth inequality and wealth aggregates. In Appendix A we also provide closed-form solutions for top shares by treating the distribution as a mixed (Generalized) Pareto distribution. Generally, the two strategies lead to identical results at the third decimal point. Figure 7 provides the results for the wealth shares of the top 1%, whereas table 1 includes other inequality measures for the raw HFCS data and the top-adjusted survey data, respectively.

What matters for the country-specific revision of wealth concentration measures is the combination of the estimated parameters of the (Generalized) Pareto distribution and the value of the transition threshold w_0. Regarding the resulting Pareto-based adjustment of wealth inequality measures, it is noteworthy that countries with the highest oversampling rates, such as Finland, France, and Portugal, experience the smallest changes in the inequal-



Note: This figure shows the change in the net wealth share of the top 1% when HFCS data are augmented with a Pareto or a Generalized Pareto tail. The resulting revisions are relatively small in countries where oversampling for the HFCS effectively targets the upper tail of the wealth distribution.

Figure 7: Share of Top 1% in Net Wealth

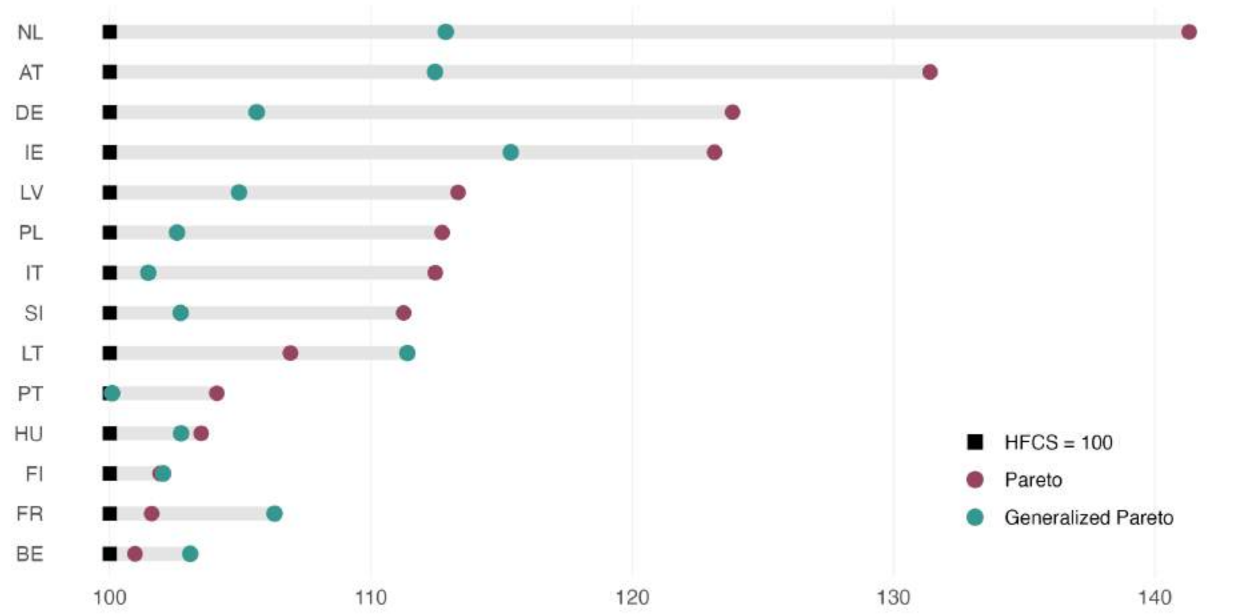
ity measures. In these countries, oversampling is based either on wealth tax data, information about the size of the primary residence, or other register-based proxies for wealth. In this regard, Germany is an exception. Top shares increase substantially with the Pareto estimation even though the effective oversampling rate of the HFCS is among the highest. The relatively substantial revisions for Germany are not surprising, as shown in Figure C.3. The regional-level oversampling implemented in Germany still results in a large gap between the HFCS and the rich list observations. The effective oversampling rate in France is similar in magnitude, but oversampling is based on administrative wealth (tax) registers (see Figure C.5). We observe considerable changes in Austria, Ireland, the Netherlands, and Lithuania. There, the top 1% shares almost double and, correspondingly, the wealth shares of the bottom 50% decrease substantially. The top 5% and 10% shares resemble the patterns of the top 1% share because the former are driven by wealth inequality within the top 1%.

Finally, we compare the corrected top 1% wealth shares with those provided in previous work using Pareto methods. In general, such a comparison is possible only to a limited extent. Prior contributions studied single countries or a small numbers of countries. Compared to the latter, we find evidence of a greater extent of wealth concentration. In line with estimates obtained using longer lists, we find major adjustments for Austria (Waltl and Chakraborty (2022): 43%; Kennickell et al. (2021) provide a variety of results ranging from 25.7% to 47.4%; Vermeulen (2018): 31-32%) and Ireland (Wildauer and Kapeller (2022): 31,7%). By contrast, we find a substantially higher share of wealth held by the top 1% in the case of the Netherlands (Vermeulen (2018): 10-19% but for a different reference year; Wildauer and Kapeller (2022): 25,8% based on a rich list of length seven as compared to our list of length 550).

In the case of the GP distribution, the revisions of the upper tail are less pronounced. This is also revealed in the CCDF plots provided in Appendix C. Due to the distribution's drift deviation from linearity, it reacts comparably more to the shape of the survey data. Consequently, the GP distribution circumvents differential under-reporting and non-response to a lesser extent than the simple Pareto distribution. Compared to raw HFCS data, the increase in the share of wealth held by the top 1% is, on average, half as large as in the Pareto estimates. There are two notable exceptions. First, we find a higher top 1% share for France than in the Pareto approach. This is due to the combination of the shape of the distribution, the effective register-based oversampling, and the long rich list. The GP approach picks up all these aspects with its flexibility. Second, GP estimates for Poland are slightly below the top shares based on raw HFCS data. Again, this is due to the flexibility of the distribution. The CCDF plot for Poland (Figure C.12) shows how the GP distribution reacts to a single survey observation that is, on the one hand, well below the bottom-ranked observation of the list and, on the other hand, way above the mass of top-ranked observations from the HFCS. On the cross-country dimension, the variation in GP-based top shares is smaller than the variation in Pareto-based shares. In sum, the Pareto-based approach is preferable over the more flexible GP approach when there is a large gap between the survey data and the rich lists, especially in combination with sparse observations at the top of the survey data.

4.3 Aggregate Wealth

Our correction for differential biases and the lack of common support between the HFCS and the ERLDB data also has implications for measures of aggregate wealth. We present the corresponding results in Figure 8 and in Table B.3 in Appendix B. Particularly in Austria and the Netherlands, the Pareto estimation increases aggregate wealth by more than 30% and 40%, respectively. In line with previous arguments, the adjustment in aggregate wealth is comparably small for countries with high effective oversampling rates. Unsurprisingly, the adjustments resulting from the GP approach are generally smaller than the Pareto-based revisions of aggregate wealth.

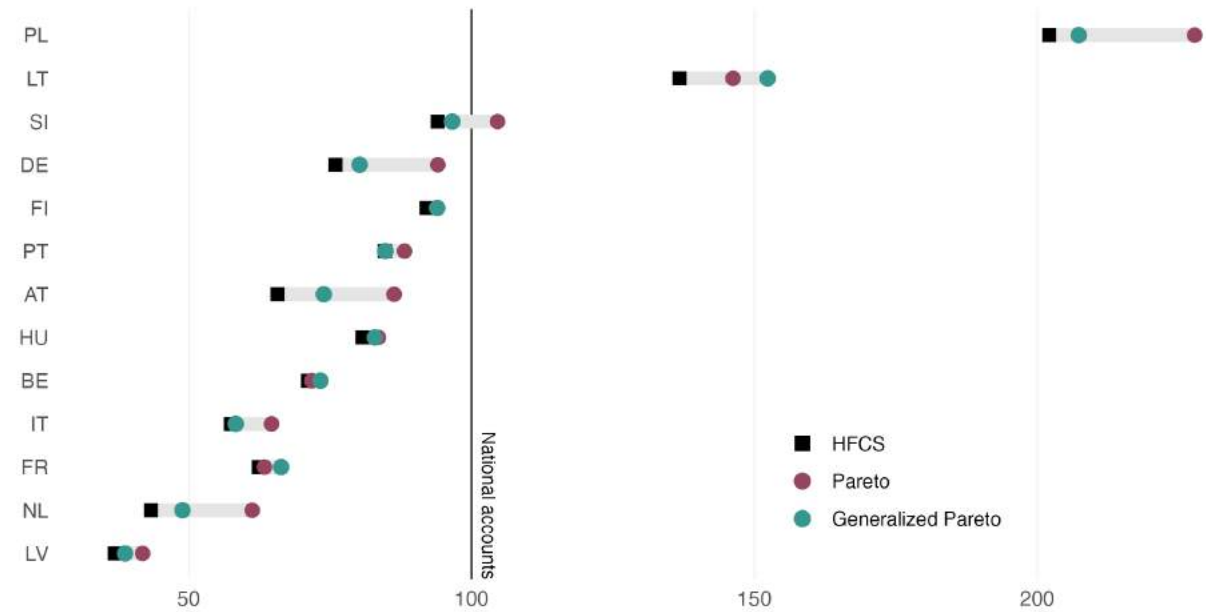


Note: This figure shows aggregate net wealth according to raw HFCS data and aggregate wealth based on the (Generalized) Pareto estimation. The aggregates based on the (Generalized) Pareto distribution are reported relative to HFCS aggregates.

Figure 8: Aggregate net wealth based on raw survey data and (Generalized) Pareto estimation

We also compare raw HFCS aggregates and top-adjusted aggregate wealth with simplified macroeconomic net wealth aggregates from National Accounts provided by Eurostat (2013). The macroeconomic accounts provide the harmonized wealth concepts, which are also the basis for implementing Distributional National Accounts (Alvaredo et al., 2020). While the macroeconomic aggregates serve as a valuable benchmark, they must be treated cautiously. Despite the underlying theoretical harmonization, the (valuation) methodologies and sectoral delimitations vary considerably across countries (Ahnert et al., 2020; Eurostat, 2021). Figure 9 provides a heterogeneous picture of the coverage ratios of macroeconomic aggregates by their aggregated microeconomic counterparts. Raw HFCS aggregates are typically well below national accounts totals, except for Poland and Lithuania. In general, some asset categories like consumer durables (furniture, cars, etc.) are excluded from national accounts, and valuables (jewelry, works of art, antiques, etc.) are included only in a few countries (Eurostat, 2013; Waltl, 2022). These assets that are missing from macroeconomic accounts but are part of the wealth concept in the HFCS, however, account for merely 5% of total real assets of the HFCS aggregates. The conceptual difference can not explain the microeconomic over-coverage for Poland and Lithuania. The ECB (Ahnert et al., 2020) thus suggests that real assets are downward-biased in the macroeconomic accounts of both countries. Our generalized regression-based method narrows the micro-macro gap for the remaining countries, particularly in Austria, Germany, and the Netherlands.

In sum, our non-discretionary regression-based approach proves to be appropriate for correcting differences in the methodological idiosyncrasies in the country-specific survey methodologies and the rich list data. In countries where wealth-correlated data is not part of the oversampling process, ex-post adjustments through Generalized (Pareto) methods based on survey data supplemented by rich lists significantly increase aggregate wealth, top shares, and other measures of inequality in case of both the Pareto and the more flexible GP distribution. The latter distribution is less suited to bridge the gap between survey data and rich lists, especially when data on the top of the distribution is sparse in the survey data. Finally, as our regression approach puts a share of trust in both survey and rich list data, it prevents over-shooting of estimates of wealth concentration and corresponding adjustments of wealth aggregates.



Note: This figure shows aggregate net wealth according to raw HFCS data and the (Generalized) Pareto estimation relative to macroeconomic aggregates from National Accounts. The macroeconomic aggregate comprises net financial assets and non-financial assets, but exclude consumer durables (furniture, cars, etc.). Most countries provide data for total fixed assets, inventories, and land. In France, the only country where all components of HFCS non-financial wealth are available in the National Accounts, these assets comprise 98% of all non-financial assets. We impute missing values for Germany, Latvia, and Portugal based on the average proportion of total fixed assets in all countries. Ireland is excluded from this figure due to the unavailability of reliable macroeconomic aggregates. For a detailed discussion and comparison of stratified HFCS aggregates and National Accounts, see e.g. Ahnert et al. (2020), Eurostat (2021) or (Waltl, 2022).

Figure 9: Aggregate Wealth Compared to Aggregates from National Accounts

Table 1: Inequality Indicators Resulting from Pareto and Generalized Pareto Estimations

	AT	BE	DE	FI	FR	HU	IE	IT	LT	LV	NL	PL	PT	SI
Gini coefficient														
HFCS	73.0	63.2	73.9	66.2	67.4	65.0	67.0	60.6	58.9	67.9	78.2	56.7	67.9	59.4
Pareto	78.8	64.0	78.7	66.8	67.9	66.2	72.6	64.6	62.7	71.6	83.6	61.0	69.1	63.1
GPareto	75.9	65.2	75.9	67.1	68.7	64.9	71.1	62.1	63.6	69.8	80.4	57.8	68.4	60.6
Share top 1%														
HFCS	22.6	16.2	18.6	14.2	17.1	20.7	14.6	12.0	14.8	19.0	20.7	14.2	23.2	15.1
Pareto	39.0	21.3	32.9	15.5	18.7	24.6	28.2	19.9	29.9	30.6	38.5	19.6	27.1	21.9
GPareto	30.7	21.2	24.6	15.2	22.0	22.3	24.3	13.3	24.8	23.2	28.5	13.5	23.8	17.0
Share top 5%														
HFCS	43.1	35.0	40.8	32.9	35.5	39.4	35.5	30.0	36.0	38.7	42.0	29.6	41.6	32.2
Pareto	57.3	39.1	52.1	32.9	36.9	42.4	47.3	37.5	46.5	48.9	57.2	36.6	45.2	39.2
GPareto	50.2	39.6	45.2	34.1	40.4	41.0	43.7	31.5	43.0	42.8	48.8	30.2	43.3	34.5
Share top 10%														
HFCS	56.4	47.2	55.4	46.8	49.2	51.4	50.0	43.4	47.9	52.1	56.6	41.3	53.9	44.0
Pareto	67.7	50.9	63.5	45.6	49.4	53.5	59.0	49.2	56.2	59.8	67.9	47.9	56.3	50.4
GPareto	61.9	51.4	58.5	47.7	52.8	52.9	56.3	44.7	54.0	55.1	61.5	42.3	55.5	46.5
Share bottom 50%														
HFCS	3.6	9.2	2.7	6.1	5.8	9.8	7.0	9.9	13.7	7.1	0.5	13.1	8.1	12.0
Pareto	2.6	5.5	4.0	9.0	8.5	9.8	6.9	10.1	NaN	4.2	2.6	11.5	8.5	10.7
GPareto	3.2	8.7	2.5	6.0	5.4	9.5	6.1	9.7	12.2	6.8	0.5	13.0	7.9	11.5
Ratio P99/P50														
HFCS	25.6	14.6	35.3	14.6	15.0	16.9	16.0	12.1	20.9	18.4	27.5	10.7	17.0	11.9
Pareto	36.0	13.0	39.1	14.7	15.2	16.5	21.1	14.4	15.6	21.3	37.8	13.6	17.6	13.9
GPareto	29.4	14.1	36.1	14.9	15.2	16.3	19.3	12.7	17.5	20.7	29.2	11.2	17.0	11.9

Note: This table shows various inequality metrics based on HFCS raw data and adjusted values from (Generalized) Pareto estimation. Countries with higher oversampling rates display smaller increases in the estimated inequality measures. The results are based on all five implicates of HFCS 2017 data.

5 Sensitivity Analysis

We perform an extensive set of sensitivity tests to stress-test our main findings. We structure this analysis along two lines. First, to address the opacities of rich lists discussed in section 2.2, we modify the ERLDB data in several dimensions. Second, we compare our baseline results to those emerging from the dominant method applied in previous work to detect the scale parameter w_{min} and the transition threshold w_0, which is an arbitrary choice of these values. Our results are highly robust to the large variety of scenarios that manipulate the ERLDB data, illuminating the advantage of our rules-based approach despite the uncertainties associated with rich list data. By contrast, we find a considerable variation in the tail adjustment across various arbitrary specifications of w_{min}.

5.1 Stability Towards Manipulations of the ERLDB

To address the uncertainty of the ERLDB data, we modify each country-specific rich list in four ways. First, we address concerns about the accuracy of the list's top end and omit absolute numbers and fractions of the top-ranked observations. We refer to the corresponding scenarios as *Drop n highest* with $n = 1, 2, 5$ and 10 and *Drop top fraction* with fraction $= 0.01, 0.05, 0.1, 0.2$ and 0.5. Second, we remove constant numbers (*Drop n lowest*) and fractions (*Drop bottom fraction*) of the bottom-ranked observations. These manipulations of the top and the bottom end of ERLDB respond to the concern that the criteria for (not) including a specific observation in a rich list are opaque (Waltl and Chakraborty, 2022; Bach et al., 2019). Third, we tackle the problem of the unclear unit of observation of each rich list. Generally, the unit of observation is certainly not homogenous as a single list may contain estimates for individuals, households, and even by (multi-generational) dynasties living in multiple households (Atkinson, 2008; Alvaredo et al., 2018; Baselgia and Martinez, 2023a; Wildauer and Kapeller, 2022). Our baseline estimates treat each rich list observation as a household. We call the scenarios that modify the observational unit *Split by n.* Specifically, we divide the wealth level of each observation by $2, 3, 4$ and 5, respectively, and generate the synthetic households. Again, we assign corresponding weight of one to each list observation.

Finally, we perform a set of sensitivity tests targeting the level of wealth reported in the lists. In the scenarios named *Vary wealth by constant*, we multiply the wealth level in the ERLDB by a constant, such as 1.2. In the scenarios *Vary wealth differentially by constant*, we increase (decrease) the wealth levels of ERLDB below a certain threshold by a constant number, and we decrease (increase) wealth levels above the threshold by a constant.[11] Table E.1 in Appendix E summarizes the sensitivity scenarios addressing the pitfalls of ERLDB. We re-estimate w_{min} and α using our generalized regression approach. We present the results of selected scenarios in Table 2 in terms of the estimated parameters of the Pareto and Generalized Pareto distribution and the top 1% wealth share. We provide the full set of results in Appendix E.1.

The results for the Pareto distribution are highly stable across the scenarios. However, we find some variation in plausible directions in the case of the most extreme scenarios. Across all countries and scenarios, the mean variation in Pareto-estimated top 1% wealth shares is less than ±3%. Correspondingly, the mean absolute change in the top 1% share is less than ±1 percentage point. In general, omitting the largest fortunes from ERLDB decreases estimated wealth concentration, while omitting the bottom-ranked observations somewhat increases wealth concentration estimates. In the latter case, we find slightly more variation across countries. Overall, the results of these sensitivity scenarios align with the intuition underlying our estimation strategy: fewer extreme observations at the very top increase α, resulting in lower top shares and vice versa. For the same reason, manipulating the observational unit as in the scenario *Split by n = 2* tends to decrease estimated wealth concentration. There are two notable exceptions from the general patterns, which are France and Italy. In the *Split by n=2* estimated wealth concentration is higher than in the baseline results, but lower in the *Drop bottom 50%* scenario.

Comparing the results between the Pareto distribution and the flexible GP distribution reveals essential insights into the relative strength of the distributions. In the case of the GP distribution, the variation across the scenarios is more pronounced. In contrast, the variation across countries is less pronounced than for the simple Pareto distribution. For

[11] Due to the similarity of the results to those from the scenarios *Vary wealth by constant*, we do not report these results here.

Table 2: Selected Sensitivity Analysis Scenarios - Manipulation of ERLDB

	AT	BE	DE	FI	FR	HU	IE	IT	LT	LV	NL	PL	PT	SI
Pareto: Alpha														
Baseline	1.32	1.61	1.40	1.89	1.73	1.51	1.47	1.64	1.38	1.41	1.33	1.63	1.47	1.57
Drop 5 highest	1.35	1.62	1.40	1.89	1.73	1.51	1.48	1.66	1.39	1.43	1.33	1.66	1.50	1.61
Drop bottom 50%	1.27	1.52	1.38	1.89	1.80	1.51	1.43	1.68	1.33	1.42	1.32	1.73	1.48	1.58
Split wealth by 2	1.30	1.69	1.44	1.88	1.52	1.50	1.53	1.54	1.40	1.45	1.36	1.64	1.48	1.65
Pareto: Share Top 1%														
Baseline	39.0	21.3	32.9	15.5	18.7	24.6	28.2	19.9	29.9	30.6	38.5	19.6	27.1	21.9
Drop 5 highest	36.1	21.0	32.7	15.4	18.6	24.5	27.7	19.6	28.8	29.2	38.2	18.7	25.6	20.5
Drop bottom 50%	43.8	24.6	34.4	15.4	17.1	24.4	30.4	18.9	33.0	30.0	39.4	17.1	26.6	21.3
Split wealth by 2	40.1	18.8	30.4	15.5	25.6	25.1	25.6	23.6	28.4	28.4	35.6	19.4	26.6	19.3
GPareto: Alpha														
Baseline	1.45	1.66	1.63	2.09	1.58	1.64	1.58	2.39	1.54	1.66	1.50	2.08	1.64	1.83
Drop 5 highest	1.53	1.73	1.67	2.15	1.66	1.68	1.78	2.53	1.59	1.70	1.56	2.14	1.68	1.89
Drop bottom 50%	1.47	1.68	1.67	2.11	1.61	1.66	1.63	2.45	1.57	1.68	1.56	2.11	1.66	1.89
Split wealth by 2	1.45	1.65	1.62	2.08	1.57	1.64	1.57	2.39	1.54	1.66	1.47	2.08	1.63	1.81
GPareto: Scale														
Baseline	175,183	207,552	218,406	201,258	224,855	54,013	499,495	201,717	37,053	27,182	173,443	63,121	133,952	100,640
Drop 5 highest	180,473	211,445	222,083	202,376	230,461	54,357	520,130	203,426	37,747	27,702	180,798	63,527	134,915	102,102
Drop bottom 50%	177,243	210,716	224,200	202,230	229,139	54,258	510,416	202,519	37,966	27,362	184,757	63,460	134,589	102,208
Split wealth by 2	174,045	205,202	214,178	199,898	221,504	53,909	490,408	201,390	36,834	27,307	160,550	62,826	133,621	98,792
GPareto: Share Top 1%														
Baseline	30.7	21.2	24.6	15.2	22.0	22.3	24.3	13.3	24.8	23.2	28.5	13.5	23.8	17.0
Drop 5 highest	27.4	19.8	23.8	14.8	20.5	21.5	20.8	12.6	23.3	22.2	27.1	13.1	22.8	16.2
Drop bottom 50%	29.8	20.7	23.9	15.1	21.5	21.8	23.4	13.0	24.0	22.5	27.1	13.3	23.3	16.3
Split wealth by 2	30.8	21.3	24.7	15.3	22.1	22.3	24.3	13.3	24.9	23.1	29.0	13.6	23.8	17.2

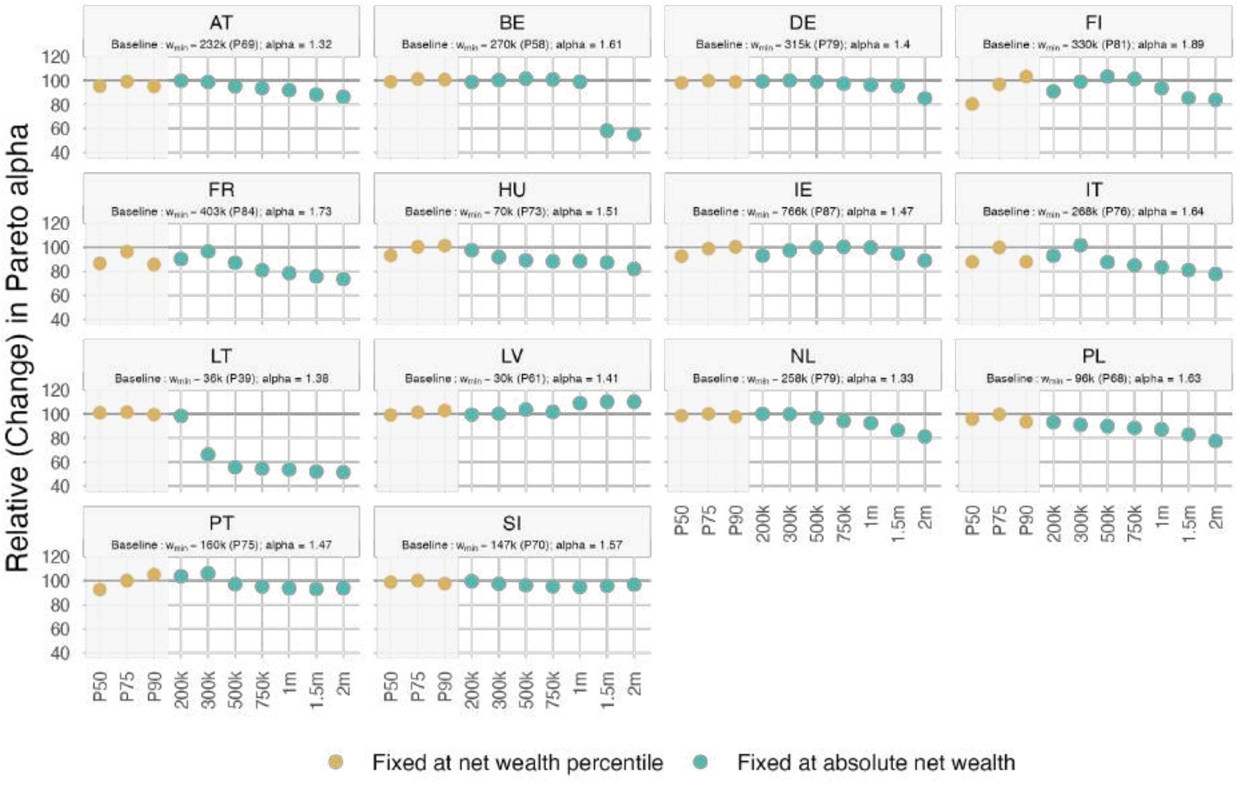
Note: This table shows the parameter estimates of the (Generalized) Pareto distribution and the share of wealth held by the top 1% for selected but stylized scenarios of the sensitivity analysis and the baseline results. The scenarios in this table are the exclusion of the top 5 observations from ERLDB (Drop 5 highest), the exclusion of the bottom 50% from ERLDB (Drop bottom 50%), and the splitting of each observation from ERLDB into two synthetic observations (Split wealth by 2). The results have been calculated on the basis of the five implicates of the HFCS 2017 using Rubin's Rule.

both distributions, omitting the top five observations from each rich list leads to lower top 1% wealth shares. However, while the *Split by n* scenarios tend to decrease estimated wealth concentration in the case of the Pareto model, they result in either no change or a slightly higher degree of wealth inequality in the case of the Generalized Pareto model. Overall, the difference to our baseline estimate of α is around the second decimal across all the *Split by n* scenarios for most countries. Relatedly, omitting the bottom 50% of rich list observations, we find little change in the estimated parameters. While the scenarios dropping bottom fractions of the list tend to result in downward revisions of top wealth shares for the GP distribution, they tend to result in upward revisions in the case of the Pareto distribution. This finding may be counter-intuitive but results from the additional flexibility of the Generalized Pareto distribution as, in our setting, this distribution puts less emphasis on rich list observations than the simpler Pareto model. By omitting a substantial fraction from a rich list, such as the bottom 50%, the upper end of the survey distribution receives even more weight in the estimation. The Pareto distribution, by contrast, reacts stronger to the rich list observations, and is better suited to bridge the non-overlapping support between survey and rich list data. We also find patterns holding for both distributions and across the scenarios, especially concerning the length of the rich list. In countries where the rich list includes relatively few observations — these are Hungary (25 observations), Italy (35 observations), and Portugal (39 observations) — the variation in the estimates across the scenarios is minimal. Shorter lists exert relatively little impact already in the baseline results. This finding matches the argument by Bach et al. (2019) and Wildauer and Kapeller (2022) that a short but country-specific list adds little to the estimation of a Pareto tail as compared to a longer list. Our findings underscore this line of reasoning, even as our method is a step forward in dealing with the survey-rich list gap.

5.2 Instability Towards Variations of w_{min}

Our second set of sensitivity tests addresses the methodology for estimating w_{min} and the replacement threshold w_0. In this analysis, we compare our baseline results to the previously dominant approach of fixing the values for both w_{min} and w_0 at arbitrary absolute values.

We first present the variation in the estimated Pareto-α across scenarios that fix w_{min} at arbitrary absolute levels (*Fix* w_{min} *at level*) and at various percentiles of the net wealth distribution (*Fix* w_{min} *at percentile*). The former set includes values typically found in previous research. Especially the comparison across fixed absolute and relative values is of interest. As our baseline results show, the optimal location parameter w_{min} varies substantially across countries regarding levels and positions. In particular for countries with low median wealth levels, high absolute values such as €500,000 or €1,000,000 are located in the top decile and might lead to inconsistent results. For each scenario, we estimate a linearized (Generalized) Pareto model given w_{min} to obtain the corresponding estimate of α and σ. Figure 10 presents the results for w_{min} set at absolute values ranging from €200,000 to €2,000,000 and at the 50th, 75th and 90th percentile of the net wealth distribution. The corresponding Figure for the GP case is Figure E.9 in Appendix E.1.



Note: This figure presents the variation in the estimated Pareto α across different location parameters (w_{min}). The location parameters are set at percentiles of the net wealth distribution and at arbitrary absolute values. Changes in α are presented relative to the baseline results with w_{min} and corresponding α calculated from the RMSE minimization of median quantile regressions.

Figure 10: Baseline Pareto α compared to arbitrary determination of w_{min}

For most countries, the variation in the estimated tail parameter α across the different scale parameters is substantial and in striking contrast to stability across the manipulations of the ERLDB. The pronounced variation of α with w_{min} translates into a substantial variation of estimates of wealth concentration. Only for three countries, Latvia, Portugal, and Slovenia, the wealth inequality estimates (and α) vary little with the choice of w_{min}.

Three conclusions emerge from our baseline results' sensitivity towards variations of w_{min}: First, when arbitrary thresholds are necessary, relative terms are preferable to absolute values, which is particularly relevant for cross-country comparisons, given the heterogeneity of wealth inequality and wealth levels. Second, despite the relative superiority of fixed percentiles over fixed net wealth levels, our rules-based approach, that considers the country-specific shape of the wealth distribution and the data quality, has to be preferred given the importance of w_{min}. Third, the location parameter generally exerts more influence on the estimated shape of the distribution than the form of the rich list.

While variations in the location parameter w_{min} directly translate into variations in estimated α, w_0 has — by design — no impact on the estimates of w_{min} and α. However, the value of w_0 affects measures of wealth concentration via the construction of the semi-parametric wealth distribution. Generally, the lower w_0, the more weight is placed on the survey data. Conditional on w_{min} and α, we hence find little variation across different and arbitrarily set values of w_0. We provide the corresponding results in Appendix E.2. For instance, the share of total wealth held by the top 1% deviates from the baseline values across various plausible values of w_0 conditional on the baseline value of w_{min} and the estimated parameters of the (Generalized) Pareto distribution only in the third to second decimal.

6 Conclusion

We provide a novel generalized regression approach to estimating heavy-tailed distributions that we apply to the distribution of wealth in 14 European countries. Much of recent research on wealth inequality, by contrast, has been centering around the U.S. and a few other countries where relevant administrative data is available. Due to substantial differences in tax legislation between countries, estimating wealth inequality based on administrative data using similar concepts for household wealth for an extensive range of countries remains an unresolved challenge. We employ data from the HFCS that provides harmonized measures of household net wealth for European countries. As with most surveys on household finances, the HFCS fails to cover the very top of the distribution due to differential survey errors along the wealth distribution, entailing biased aggregate wealth and wealth concentration estimates. We hence supplement the HFCS with rich lists that provide, to date, the most comprehensive data source on the wealth held by the ultra-wealthy, and we introduce the first systematic compilation of rich lists in the European Rich List Database (ERLDB). ERLDB is also the first database that includes country-specific lists for more than one country. Combining the HFCS with the ERLDB, we can provide novel measures of aggregate wealth and wealth inequality for 14 countries based on a (Generalized) Pareto estimation framework that uses country-specific lists and a cross-country harmonized concept of wealth. Such measures are direly needed. For example, the World Inequality Database (WID) publishes wealth inequality statistics for almost all countries around the globe. However, for the vast majority of countries, these measures are imputed based on estimates of income inequality and the cross-country correlation of income and wealth inequality among the few countries for which both estimates are available (Bajard et al., 2022).

Our generalized regression approach to estimating the (Generalized) Pareto distribution accounts for differential non-response and under-reporting of wealth in the survey data. Linearization of the cumulative density function allows for the intuitive but robust median quantile regression approach as our preferred estimation technique, with the location parameter, survey weight correction, and simulation thresholds derived from the distribution's stochastic definition of regression results. Our approach circumvents visual inspection of distributions and discretionary decisions, and addresses heterogeneities in wealth accumulation, inequality, and idiosyncrasies in the underlying data. It is hence easily applicable to other countries and periods. From this perspective, our method is particularly relevant for estimating top wealth and top income shares and implementing Distributional National and Financial Accounts.

Compared to unadjusted survey data, our correction for differential non-response and under-reporting results in a substantial revision of aggregate wealth and wealth concentration measures. In the two extreme cases of the Netherlands and Austria, the top 1% wealth share almost doubles to 38.5% and 39.0%, respectively. By contrast, the revision of inequality measures is less pronounced for countries where differential errors are less extreme, especially in France and Finland. In these countries, the HFCS uses administrative data to oversample wealthy households. Accordingly, we find a significant negative correlation between the effective oversampling rate of the top tail in the HFCS and the stability of inequality metrics across the raw survey data and the (Generalized) Pareto distribution-based tail adjustments. The tail adjustments also translate into revisions of aggregate wealth, ranging from only 2% or 3% in France, Finland, and Belgium to almost 40% in the Netherlands and Austria.

Prior work cautions against using rich lists in Pareto-based estimations of wealth inequality. This cautionary tale, to some extent, stems from rich list data taken at face value. For instance, Kopczuk and Saez (2004) and Alvaredo et al. (2018) compare rich list-based top shares to mortality multiplier-based estimates. The former overshoot the latter substantially, and the implied Pareto distributions are hard to reconcile. More fundamentally, the difference is so striking that the question of whether the Pareto estimates obtained from the rich list and the mortality-multiplier approach describe the same population. Our quantile regression approach circumvents over-shooting by using data from rich lists jointly with survey data and bridging the gap (the lack of common support) between these sources by putting a share of trust in either source. Neither survey data nor rich list data are taken at face value. Using our median quantile regression approach, we find stable tail adjustments towards a large variety of sensitivity scenarios that manipulate the ERLDB data. By contrast, our results vary substantially across different arbitrary fixed location paramter w_{min}.

We observe, by contrast, a substantial variation across the previously dominant (arbitrary) specification of w_{min}. We conclude that the estimation method is more important than the quality of the rich list.

While our main contribution is methodological, the results have important policy implications. We stress two of them. First, improving the estimation of wealth aggregates and wealth inequality is key to advancing the design and evaluation of wealth taxes, a discussion that has recently gained momentum (Saez and Zucman, 2019; Bastani and Waldenström, 2020;

Scheuer and Slemrod, 2021; Advani et al., 2021a; Advani et al., 2021b; Adam and Miller, 2021). While the number of countries levying recurrent net wealth taxes has decreased since the 1990s, some countries expressed continued interest in wealth taxation (OECD, 2018). Biased estimates of wealth aggregates and wealth inequality entail biased expectations of potential tax revenue, the redistribute effect of wealth taxes, and of behavioral and real responses to wealth taxation. Due to the typical high exemption thresholds of wealth and estate taxes (Scheuer and Slemrod, 2021), an accurate measurement of the top of the wealth distribution is crucial. Related, the Pareto tail parameter we estimate is an essential ingredient of optimal tax formulas (see, for example, the sufficient statistics approach to the taxation of capital by Saez and Stantcheva, 2018). In sum, this paper also informs the discussion on the revenue potential and distributional implications of wealth taxes because relevant administrative data is not available for most of the countries included in our sample. Second, we provide cross-country comparable measures of aggregate wealth and wealth inequality that are generally revised upwards compared to raw survey data. Evidence for the U.S. and Australia shows that people tend to underestimate actual levels of wealth inequality (Hauser and Norton, 2017; Norton et al., 2014; Norton and Ariely, 2011). Such inequality perceptions are even more pronounced once the *missing rich* are taken into account.

References

Acciari, P. and S. Morelli (2022). Wealth Transfers and Net Wealth at Death: Evidence from the Italian Inheritance Tax Records 1995-2016. *Measuring Distribution and Mobility of Income and Wealth*. Ed. by R. Chetty, J. M. Friedman, J. C. Gornick, and A. B. Kennickell.

Adam, S. and H. Miller (2021). The Economic Arguments For and Against a Wealth Tax. *Fiscal Studies* 42 (3-4), 457–483. DOI: 10.1111/1475-5890.12288.

Advani, A., G. Bangham, and J. Leslie (2021a). The UK's Wealth Distribution and Characteristics of High-Wealth Households. *Fiscal Studies* 42 (3-4), 397–430. DOI: 10.1111/1475-5890.12286.

Advani, A., A. Summers, and H. Tarrant (2022). Who are the Super-Rich? The Wealth and Connections of the Sunday Times Rich List. *CAGE Warwick Policy Briefing* 37.

Advani, A. and H. Tarrant (2021). Behavioural Response to a Wealth Tax. *Fiscal Studies* 42 (3-4), 509–538. DOI: 10.1111/1475-5890.12283.

Advani, A., H. Hughson, and H. Tarrant (2021b). Revenue and Distributional Modelling for a UK Wealth Tax. *Fiscal Studies* 42 (3-4), 699–736. DOI: 10.1111/1475-5890.12280.

Ahnert, H., I. K. Kavonius, J. Honkkila, and P. Sola (2020). Understanding Household Wealth: Linking Macro and Micro data to Produce Distributional Financial Accounts. *European Central Bank Statistics Paper Series* 37.

Albers, T. N., C. Bartels, and M. Schularick (2022). Wealth and Its Distribution in Germany. *CESifo Working Paper* 9739.

Alvaredo, F., A. B. Atkinson, L. Chancel, T. Piketty, E. Saez, and G. Zucman (2020). Distributional National Accounts Guidelines. Methods and Concepts used in the World Inequality Database. *World Inequality Lab*.

Alvaredo, F., A. B. Atkinson, T. Piketty, and E. Saez (2013). The Top 1 Percent in International and Historical Perspective. *Journal of Economic Perspectives* 27 (3), 3–20. DOI: 10.1257/jep.27.3.3.

Alvaredo, F., A. B. Atkinson, and S. Morelli (2018). Top Wealth Shares in the UK over More than a Century. *Journal of Public Economics* 162, 26–47. DOI: 10.1016/j.jpubeco.2018.02.008.

Atkinson, A. B. (2008). Wealth Concentration Among the Rich. *Personal wealth from a global perspective*. Ed. by J. B. Davies. UNU-WIDER Studies in Development Economics. Oxford: Oxford University Press, 64–89.

Atkinson, A. B. (2017). Pareto and the Upper Tail of the Income Distribution in the UK: 1799 to the Present. *Economica* 84 (334), 129–156. DOI: 10.1111/ecca.12214.

Atkinson, A. B. and A. J. Harrison (1978). *Distribution of Personal Wealth in Britain*. Cambridge, MA: Cambridge University Press.

Atkinson, A. B. and T. Piketty (2007). *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*. Oxford University Press.

Bach, S., A. Thiemann, and A. Zucco (2019). Looking for the Missing Rich: Tracing the Top Tail of the Wealth Distribution. *International Tax and Public Finance* 26 (6), 1234–1258. DOI: 10.1007/s10797-019-09578-1.

Bajard, F., L. Chancel, R. Moshrif, and T. Piketty (2022). *Global Wealth Inequlity on WID.world: Estimtes and imputations*. World Inequality Lab - Technical Note 2021/06. Paris: World Inequality Lab.

Baselgia, E. and I. Martinez (2023a). Behavioral Responses to Special Tax Regimes for the Super-Rich: Insights from Swiss Rich Lists. *EU Tax Observatory Working Paper* 12.

— (2023b). *Using Rich Lists to Study the Super-Rich and Top Wealth Inequality: Insights from Switzerland*. https://www.dropbox.com/s/zxfgwmb4zp5szu4/superrich_descriptive_v1.pdf?dl=0.

Bastani, S. and D. Waldenström (2020). How Should Capital be Taxed? *Journal of Economic Surveys* 34 (4), 812–846. DOI: 10.1111/joes.12380.

Batty, M., J. Bricker, J. Briggs, S. Friedman, D. Nemschoff, E. Nielsen, K. Sommer, and A. H. Volz (2021). The Distributional Financial Accounts of the United States. Chetty, R., J. N. Friedman, J. C. Gornick, B. Johnson, and A. Kennickell. *Measuring Distribution and Mobility of Income and Wealth*. University of Chicago Press.

Benhabib, J. and A. Bisin (2018). Skewed Wealth Distributions: Theory and Empirics. *Journal of Economic Literature* 56 (4), 1261–1291. DOI: 10.1257/jel.20161390.

Benhabib, J., A. Bisin, and S. Zhu (2011). The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents. *Econometrica* 79 (1), 123–157. DOI: 10.3982/ECTA8416.

— (2015). The Wealth Distribution in Bewley Cconomies with Capital Income Risk. *Journal of Economic Theory* 159, 489–515. DOI: 10.1016/j.jet.2015.07.013.

— (2016). The Distribution of Wealth in the Blanchard-Yaari Model. *Macroeconomic Dynamics* 20 (2), 466–481. DOI: 10.1017/S1365100514000066.

Berman, Y. and S. Morelli (2022). On the Distribution of Estates and the Distribution of Wealth: Evidence from the Dead. *Measuring Distribution and Mobility of Income and Wealth*. Ed. by R. Chetty, J. M. Friedman, J. C. Gornick, and A. B. Kennickell.

Blanchet, T., I. Flores, and M. Morgan (2022). The Weight of the Rich: Improving Surveys using Tax Data. *The Journal of Economic Inequality* 20 (1). DOI: 10.1007/s10888-021-09509-3.

Blanchet, T., J. Fournier, and T. Piketty (2021). Generalized Pareto Curves: Theory and Applications. *Review of Income and Wealth* 68 (1), 263–288. DOI: 10.1111/roiw.12510.

Blanchet, T., B. Garbinti, J. Goupille, and C. Martìnez-Toledano (2018). Applying Generalized Pareto Curves to Inequality Analysis. *American Economic Association: Papers & Proceedings* 108, 114–118. DOI: 10.1257/pandp.20181075.

Bricker, J., A. Henriques, J. Krimmel, and J. Sabelhaus (2016). Measuring Income and Wealth at the Top Using Administrative and Survey Data. *Brookings Papers on Economic Activity* 47 (1), 261–331.

Brzezinski, M. (2014). Do Wealth Distributions Follow Power Laws? Evidence from "Rich Lists". *Physica A: Statistical Mechanics and its Applications* 406, 155–162. DOI: 10.1016/j.physa.2014.03.052.

Brzezinski, M., K. Sałach, and M. Wroński (2020). Wealth Inequality in Central and Eastern Europe: Evidence from Household Survey and Rich Lists' Data Combined. *Economics of Transition and Institutional Change* 28 (4), 637–660. DOI: 10.1111/ecot.12257.

Campolieti, M. (2018). Heavy-Tailed Distributions and the Distribution of Wealth: Evidence from Rich Lists in Canada, 1999–2017. *Phisica A: Statistical Mechanics and its Applications* 50 (3), 263–272. DOI: 10.1016/j.physa.2018.02.057.

Champernowne, D. G. (1953). A Model of Income Distribution. *The Economic Journal* 63 (250), 318–351. DOI: 10.2307/2227127.

Charpentier, A. and E. Flachaire (2022). Pareto Models for Top Income and Wealth. *The Journal of Economic Inequality* 20 (1), 1–25. DOI: 10.1007/s10888-021-09514-6.

Chatterjee, A., C. Léo, and A. Gethin (2022). Wealth Inequality in South Africa, 1993–2017. *The World Bank Economic Review* 36 (1), 19–36. DOI: 10.1093/wber/lhab012.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-Law Distributions in Empirical Data. *SIAM Review* 51 (4), 661–703. DOI: 10.1137/070710111.

Cowell, F. A. (2011). *Measuring Inequality.* 3rd ed. LSE Perspectives in Economic Analysis. Oxford: Oxford University Press.

D'Alessio, G. and I. Faiella (2002). Non-Response Behaviour in the Bank of Italy's Survey of Household Income and Wealth. *Banca D'Italia Working Paper* 462.

Dalitz, C. (2016). Estimating Wealth Distribution: Top Tail and Inequality. *Hochschule Niederrhein, Fachbereich Elektrotechnik & Informatik Working Paper* 2016-01.

Davies, J. B. and A. E. Shorrocks (2000). The Distribution of Wealth. *Handbook of Income Distribution Volume 1.* Ed. by A. B. Atkinson and F. Bourguignon. Elsevier, 605–675. DOI: 10.1016/S1574-0056(00)80014-7.

Eckerstorfer, P., J. Halak, J. Kapeller, B. Schütz, F. Springholz, and R. Wildauer (2016). Correcting for the Missing Rich: An Application to Wealth Survey Data. *Review of Income and Wealth* 62 (4), 605–627. DOI: 10.1111/roiw.12188.

Engel, J., P. G. Riera, J. Grilli, and P. Sola (2022). Developing Reconciled Quarterly Distributional National Wealth Accounts – Insight into Inequality and Wealth Structures. *European Central Bank Working Paper Series* 2687.

European Central Bank (2020). The Eurosystem Household Finance and Consumption Survey: Methodological Report for the 2017 Wave. *ECB Statistics Paper Series* 35.

Eurostat (2013). *European System of Accounts: ESA 2010.* EDC collection. Luxembourg: Publications Office of the European Union.

— (2021). *Meta Data on the Collection of Non-Financial Assets.* https://ec.europa.eu/eurostat/web/national-accounts/methodology/member-states-accounts/non-financial-assets.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2016). Heterogeneity in Returns to Wealth and the Measurement of Wealth Inequality. *American Economic Review* 106 (5), 651–55. DOI: 10.1257/aer.p20161022.

Flachaire, E., N. Lustig, and A. Vigorito (2021). *Underreporting of Top Incomes and Inequality: An Assessment of Correction Methods using Linked Survey and Tax Data.* Paper prepared for the 36th IARIW Virtual General Conference. Norway.

Föllmi, R. and I. Martinez (2017). Volatile Top Income Shares in Switzerland? Reassessing the Evolution Between 1981 and 2010. *The Review of Economics and Statistics* 99 (5), 793–809. DOI: 10.1162/REST_a_00644.

Gabaix, X. (1999). Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics* 114 (3), 739–767. DOI: 10.1162/003355399556133.

— (2016). Power Laws in Economics: An Introduction. *Journal of Economic Perspectives* 30 (1), 185–206. DOI: 10.1257/jep.30.1.185.

Gabaix, X. and R. Ibragimov (2011). Rank -1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business Economics and Statistics* 29 (1), 24–39. DOI: 10.1198/jbes.2009.06157.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The Dynamics of Inequality. *Econometrica* 84 (6), 2071–2111. DOI: 10.3982/ECTA13569.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2021). Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France. *Journal of the European Economic Association* 19 (1), 620–663. DOI: 10.1093/jeea/jvaa025.

Hauser, O. P. and M. I. Norton (2017). (Mis)perceptions of Inequality. *Current Opinion in Psychology* 18, 21–25. DOI: 10.1016/j.copsyc.2017.07.024.

Hlasny, V. and P. Verme (2018). Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods using the EU SILC Data Vladimir. *Econometrics* 6 (30), 1–38. DOI: 10.3390/econometrics6020030.

Iacono, R. and E. Palagi (2023). A Micro-Perspective on r > g. *Economica* 90 (358), 531–556. DOI: doi.org/10.1111/ecca.12464.

Jakobsen, K., K. Jakobsen, H. Kleven, and G. Zucman (2020). Wealth Taxation and Wealth Accumulation: Theory and Evidence from Denmark. *The Quarterly Journal of Economics* 135 (1), 329–388. DOI: 10.1093/qje/qjz032.

Jenkins, S. P. (2017). Pareto Models, Top Incomes and Recen Trends in UK Income Inequality. *Economica* 84 (334), 261–289. DOI: 10.1111/ecca.12217.

Jones, C. (2015). Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality. *Journal of Economic Perspectives* 29 (1), 29–46. DOI: 10.1257/jep.29.1.29.

Jones, C. and J. Kim (2018). A Schumpeterian Model of Top Income Inequality. *Journal of Political Economy* 126 (5), 1785–1826. DOI: 10.1086/699190.

Kapeller, J., R. Wildauer, and S. Leitch (2021). A European Wealth Tax for a Fair and Green Recovery. GPERC81.

Kaplan, S. N. and J. D. Rauh (2013). Family, Education, and Sources of Wealth among the Richest Americans, 1982-2012. *American Economic Review* 103 (3), 158–62. DOI: 10.1257/aer.103.3.158.

Kennickell, A. B. (2021). Chasing the Tail: A Generalized Pareto Distribution Approach to Estimating Wealth Inequality. *Stone Center Working Paper Series* 37.

— (2003). A Rolling Tide: Changes in the Distribution of Wealth in the U.S., 1989–2001. *The Levy Economics Institute Working Paper* 393.

— (2008). The Role of Over-Sampling of the Wealthy in the Survey of Consumer Finances. *Irving Fisher Committee Bulletin* 28.

— (2019). The Tail That Wags: Differences in Effective Right Tail Coverage and Estimates of Wealth Inequality. *The Journal of Economic Inequality* 17 (4), 443–459. DOI: 10.1007/s10888-019-09424-8.

Kennickell, A. B., P. Lindner, and M. Schürz (2021). A New Instrument to Measure Wealth Inequality: Distributional Wealth Accounts. *Monetary Policy and the Economy* 2021 (Q4), 61–85.

Kennickell, A. B. and R. L. Woodburn (1999). Consistent Weight Design for the 1989, 1992, and 1995 SCFs, and the Distribution of Wealth. *Review of Income and Wealth* 45 (2), 193–215. DOI: 10.1111/j.1475-4991.1999.tb00328.x.

Klass, O. S., O. Biham, M. Levy, O. Malcai, and S. Solomon (2006). The Forbes 400 and the Pareto Wealth Distribution. *Economics Letters* 90 (2), 290–295. DOI: 10.1016/j.econlet.2005.08.020.

Koenker, R. and G. Bassett (1978). Regression Quantiles. *Econometrica* 46 (1), 33–50. DOI: 10.2307/1913643.

Kopczuk, W. (2015). What do We Know About the Evolution of Top Wealth Shares in the United States? *Journal of Economic Perspectives* 29 (1), 47–66. DOI: 10.1257/jep.29.1.47.

Kopczuk, W. and E. Saez (2004). Top Wealth Shares in the United States, 1916–2000: Evidence from Estate Tax Returns. *National Tax Journal* 57 (2), 445–487. DOI: 10.17310/ntj.2004.2S.05.

Kuznets, S. (1953). Shares of Upper Income Groups in Income and Saving. *Shares of Upper Income Groups in Income and Saving*. Ed. by S. Kuznets and E. Jenks. National Bureau of Economic Research, 171–218.

Langousis, A., A. Mamalakis, M. Puliga, and R. Deidda (2016). Threshold Detection for the Generalized Pareto Distribution: Review of Representative Methods and Application to the NOAA NCDC Daily Rainfall Database. *Water Resources Research* 52 (4), 2659–2681. DOI: 10.1002/2015WR018502.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. 2nd ed. John Wiley and Sons. DOI: 10.1002/9781119013563.

Lundberg, J. and D. Waldenström (2018). Wealth Inequality in Sweden: What Can We Learn from Capitalized Income Tax Data? *Review of Income and Wealth* 64 (3), 517–541. DOI: 10.1111/roiw.12294.

Luo, C. and G. Chen (2021). Re-Estimating National Wealth Inequality with Incorporating the Rich Lists in China. *China Economic Quarterly International* 1 (4), 295–307. DOI: doi.org/10.1016/j.ceqi.2021.11.001.

Lustig, N. (2020). The "Missing Rich" in Household Surveys: Causes and Correction Approaches. *ECINEQ Working Paper* 520.

Luttmer, E. G. (2007). Selection, Growth, and the Size Distribution of Firms. *The Quarterly Journal of Economics* 122 (3), 1103–1144. DOI: 10.1162/qjec.122.3.1103.

Martínez-Toledano, C. (2022). "House Price Cycles, Wealth Inequality and Portfolio Reshuffling."

Moretti, E. and D. J. Wilson (2022). Taxing Billionaires: Estate Taxes and the Geographical Location of the Ultra-Wealthy. *American Economic Journal: Economic Policy, forthcoming*.

Nirei, M. (2009). *Pareto Distributions in Economic Growth Models*. IIR Working Paper 09-05.

Nirei, M. and S. Aoki (2016). Pareto Distribution of Income in Neoclassical Growth Models. *Review of Economic Dynamics* 20, 25–42. DOI: 10.1016/j.red.2015.11.002.

Norton, M. I. and D. Ariely (2011). Building a Better America – One Wealth Quintile at a Time. *Perspectives on Psychological Science* 6 (1), 9–12. DOI: 10.1177/1745691610393524.

Norton, M. I., D. T. Neal, C. L. Govan, D. Ariely, and E. Holland (2014). The Not-So-Common-Wealth of Australia: Evidence for a Cross-Cultural Desire for a More Equal Distribution of Wealth. *Analyses of Social Issues and Public Policy* 14 (1), 339–351. DOI: 10.1111/asap.12058.

OECD (2018). *The Role and Design of Net Wealth Taxes in the OECD*. OECD Tax Policy Studies 26. DOI: 10.1787/9789264290303-en.

Osier, G. (2016). *Unit Non-Response in Household Wealth Surveys: Experience from the Eurosystem's Household Finance and Consumption Survey*. ECB Statistics Paper 15. European Central Bank (ECB), Frankfurt a. M.

Pareto, V. (1965). La Courbe de la Repartition de la Richesse. *Oevres Completes de Vilfredo Pareto*. Ed. by G. Busino. Vol. 3. Librairie Droz, 1–15.

Perret, S. (2021). Why were Most Wealth Taxes Abondoned and is This Time Different? *Fiscal Studies* 42 (3-4), 539–563. DOI: 10.1111/1475-5890.12278.

Pfeffer, F. T., R. F. Schoeni, A. B. Kennickell, and P. Andreski (2016). Measuring Wealth and Wealth Inequality: Comparing two U.S. Surveys. *Journal of Economic and Social Measurement* 41 (2), 103–120. DOI: 10.3233/JEM-160421.

Piketty, T. (2003). Income Inequality in France, 1901-1998. *Journal of Political Economy* 111 (5), 1004–1042. DOI: 10.1086/376955.

Piketty, T., G. Postel-Vinay, and J.-L. Rosenthal (2006). Wealth Concentration in a Developing Economy: Paris and France, 1807–1994. *American Economic Review* 96 (1), 237–256. DOI: 10.1257/000282806776157614.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics* 118 (1), 1–41. DOI: 10.1162/00335530360535135.

Piketty, T., E. Saez, and G. Zucman (2022). Twenty Years and Counting: Thoughts about Measuring the Tpper Tail. *The Journal of Economic Ineqaulity* 20 (1), 255–264. DOI: 10.1007/s10888-022-09536-8.

Piketty, T. and G. Zucman (2015). Wealth and Inheritance in the Long Run. *Handbook of Income Distribution*. Ed. by A. B. Atkinson and F. Bourguignon. Vol. 2. Elsevier, 1303–1368. DOI: B978-0-444-59429-7.00016-9.

Ravallion, M. (2022). Missing Top Income Recipients. *The Journal of Economic Inequality* 20 (1), 205–222. DOI: 10.1007/s10888-022-09530-0.

Roine, J. and D. Waldenström (2009). Wealth Concentration over the Path of Development: Sweden, 1873–2006. *The Scandinavian Journal of Economics* 111 (1), 151–187. DOI: 10.1111/j.1467-9442.2008.01558.x.

— (2015). Long-Run Trends in the Distribution of Income and Wealth. *Handbook of Income Distribution*. Ed. by A. B. Atkinson and F. Bourguignon. Vol. 2. Amsterdam: North-Holland, 469–592. DOI: 10.1016/B978-0-444-59428-0.00008-4.

Saez, E. and S. Stantcheva (2018). A Simpler Theory of Optimal Capital Taxation. *Journal of Public Economics* 162, 120–142. DOI: 10.1016/j.jpubeco.2017.10.004.

Saez, E. and G. Zucman (2016). Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data. *The Quarterly Journal of Economics* 131 (2), 519–578. DOI: 10.1093/qje/qjw004.

— (2019). Progressive Wealth Taxation. *Brookings Papers on Economic Activity* Fall 2019, 437–511.

— (2020a). The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts. *Journal of Economic Perspectives* 34 (4), 3–26. DOI: 10.1257/jep.34.4.3.

— (2020b). *Trends in US Income and Wealth Inequality: Revising after the Revisionists*. NBER Working Paper 27921. DOI: 10.3386/w27921.

— (2022). Comments on Smith, Zidar and Zwick (2021).

Salach, K. and M. Brzezinski (2020). Political Connections and the Super-Rich in Poland. *Univeresity of Warsaw Working Paper* 17/2020. DOI: 10.2139/ssrn.3631195.

Scheuer, F. and J. Slemrod (2020). Taxation and the Superrich. *Annual Review of Economics* 12, 189–211. DOI: 10.1146/annurev-economics-081919-115106.

— (2021). Taxing our Wealth. *Journal of Economic Perspectives* 35 (1), 207–230. DOI: 10.1257/jep.35.1.207.

Schulter, C. (2020). On Zipf's Law and the Bias of Zipf Regressions. *Empirical Economics* 61, 529–548. DOI: 10.1007/s00181-020-01879-3.

Schulz, J. and M. Milaković (2021). How Wealthy are the Rich? *Review of Income and Wealth, forthcoming*. DOI: 10.1111/roiw.12550.

Seim, D. (2017). Behavioral Response to Wealth Taxes: Evidence from Sweden. *American Economic Journal: Economic Policy* 9 (4), 395–421. DOI: 10.1257/pol.20150290.

Sheather, S. J. and C. M. Jones (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 53 (3), 683–690. DOI: 10.1111/j.2517-6161.1991.tb01857.x.

Smith, M., O. Zidar, and E. Zwick (2021). Response to Saez and Zucman (2020)'s "Comments on Smith Zidar Zwick (2020)".

Smith, M., O. M. Zidar, and E. Zwick (2023). Top Wealth in America: New Estimates under Heterogeneous Returns. *Quarterly Journal of Economics* 138 (1), 515–573. DOI: 10.1093/qje/qjac033.

Summers, A. (2021). Ways of Taxing Wealth: Alternatives and Interactions. *Fiscal Studies* 42 (3-4), 485–507. DOI: 10.1111/1475-5890.12285.

Tisch, D. and E. Ischinsky (2023). Top Wealth and Its Historical Origins: An Analysis of Germany's Largest Privately Held Fortunes in 2019. *Max Planck Institute for the Study of Society Working Paper* 23/1.

Vermeulen, P. (2016). Estimating the Top Tail of the Wealth Distribution. *American Economic Review* 106 (5), 646–650. DOI: 10.1257/aer.p20161021.

— (2018). How Fat is the Top Tail of the Wealth Distribution? *Review of Income and Wealth* 64 (2), 357–387. DOI: 10.1111/roiw.12279.

Waltl, S. R. and R. Chakraborty (2022). Missing the Wealthy in the HFCS: Micro Problems with Macro Implications. *The Journal of Economic Inequality* 20 (1), 169–203. DOI: 10.1007/s10888-021-09519-1.

Waltl, S. R. (2022). Wealth Inequality: A Hybrid Approach Toward Multidimensional Distributional National Accounts in Europe. *Review of Income and Wealth* 68 (1), 74–108. DOI: 10.1111/roiw.12519.

Wildauer, R. and J. Kapeller (2022). Tracing the Invisible Rich: A New Approach to Modelling Pareto Tails in Survey Data. *Labour Economics* 75. DOI: 10.1016/j.labeco.2022.102145.

Wold, H. O. A. and P. Whittle (1957). A Model Explaining the Pareto Distribution of Wealth. *Econometrica* 25 (4), 591–595. DOI: 10.2307/1905385.

Zucman, G. (2019). Global Wealth Inequality. *Annual Review of Economics* 11, 109–138. DOI: 10.1146/annurev-economics-080218-025852.

Zwijnenburg, J. (2022). The Use of Distributional National Accounts in Better Capturing the Top Tail of the Distribution. *The Journal of Economic Inequality* 20 (1), 245–254. DOI: 10.1007/s10888-022-09534-w.

Appendix - Contents

Appendix A. Technical Appendix.

Appendix B. Supporting Material: Data and Main Results.

Appendix C. CCDF Plots by Country and Implicate

Appendix D. Transition Threshold Parameter by Country and Implicate.

Appendix E. Sensitivity Analysis.

Appendix E.1. Sensitivity Analysis: ERLDB and w_{min}.

Appendix E.2. Sensitivity Analysis: w_0.

A Technical appendix

In this appendix, we describe the two approaches for estimating top-corrected measures of aggregate wealth and wealth inequality. The first approach simulates new top-tail observations based on the parameter estimates of the (Generalized) Pareto distribution and combines them with non-tail observations from the HFCS. This strategy is our preferred approach because we can treat the resulting combination of non-tail observations from the HFCS and parameter-based tail observation as a top-corrected data set spanning the entire range of net wealth. The second approach expresses the wealth distribution as a weighted sum of conditional means. We explain the intuition behind this strategy for the case of top wealth shares. Following the same logic, one can derive expressions for other distributional measures and aggregates.

A.1 Simulating the Tail

This approach to obtain a top-corrected wealth distribution relies on the parameter estimates of the (Generalized) Pareto distribution and the definition of the tail length as presented in section 3. We simulate new observations within the tail and combine the top-corrected tail with non-tail observations from the HFCS. Overall, we obtain a distribution based on HFCS data for $w_i \leq w_0$, w_0 is again the transition threshold parameter, and the parameter estimates of the parametric distribution beyond w_0. For given (Generalized) Pareto parameters w_{min} and α we can determine the theoretical value of the net wealth of any observations with rank $i \in [1, \sum_{w_i \geq w_{min}} n(w_i)]$. Note that $n(w_i)$ denotes the number of observations with value w_i, i.e. the sum of the (survey) weights and that for a known transition value $w_0 > w_{min}$, $\sum_{w_i \geq w_{min}} n(w_i) = \frac{1}{F(w_0)} \sum_{w_i \in [w_{min}, w_0]} n(w_i)$.

The simulation approach relies on the definition of the complementary cumulative density function (CCDF), [12] as $1 - F(w_i)$, which gives the fraction of observations with net wealth equal or larger than w_i. With observations ranked in descending order, such that rank $i = 1$ corresponds to the observation with the largest w_i, the CCDF is equivalent to $\frac{i}{\sum_{w_i \geq w_{min}} n(w_i)}$ such that

$$\text{CCDF} = \frac{i}{\sum_{w_i \geq w_{min}} n(w_i)} = \frac{i}{\frac{1}{F(w_0)} \sum_{w_i \in [w_{min}, w_0]} n(w_i)} \tag{16}$$

In the case of the simple Pareto distribution, wealth levels of the simulated observations are hence given

[12] The CCDF of the Pareto distribution is given by $(\frac{w_i}{w_{min}})^{-\alpha}$ while the CCDF of the Generalized Pareto distribution is given by $(1 + \xi \frac{w_i - w_{min}}{\sigma})^{-\frac{1}{\xi}}$.

by

$$\text{CCDF (Pareto)} = \left(\frac{w_i}{w_{min}}\right)^{-\alpha}$$
$$w_i = w_{min}\left(\frac{i}{\frac{1}{F(w_0)}\sum_{w_i\in[w_{min},w_0]} n(w_i)}\right)^{-1/\alpha} \tag{17}$$

Given the tail length, i.e. the number of households with $w_i \leq w_0$, we simulate the corresponding number of wealth levels and assign a uniform weight of 1 to each observation.

A.1.1 Deriving Top Shares from Estimated Parameters

In the second approach, we obtain a top-corrected wealth distribution as weighted conditional mean, following the approach proposed by Charpentier and Flachaire (2022). We explain this for the case of top wealth shares. A top wealth share is the share of aggregate net wealth held by households in a top percentile, e.g. the top 1% share is the share of wealth held by the richest 1% of households. In discussing Pareto models for top incomes, Charpentier and Flachaire (2022) propose expressing top shares as a ratio of sums, or in the case of a mixed distribution, as a weighted ratio of conditional means. In a mixed distribution containing an empirical lower part and a parametric upper tail, fully separated at some threshold value x_{min}, let p be the percentile of x_{min} in the mixed distribution, q an arbitrary percentile in the mixed distribution, and r the corresponding percentile in either component distribution. Then, if $q > p$, the rth percentile in the parametric tail corresponds to the qth percentile in the mixed distribution.

$$TS_{Q,q} = \frac{\sum_{x_i>Q(X,r)} x_i n(x_i)}{\sum_{x_i<x_{min}} x_i n(x_i) + \sum_{x_i\geq x_{min}} x_i n(x_i)} \tag{18}$$
$$= \frac{(1-q)E[\bar{X} \mid X \geq Q(X,r)]}{pE[\bar{X} \mid X < x_{min}] + (1-p)E[\bar{X} \mid X \geq x_{min}]} \tag{19}$$

Since x_{min} separates the bottom (non-tail) and top (parametric tail) distributions, r can be derived from p and q.

$$r = \begin{cases} q > p & r = \frac{q-p}{1-p} \\ q = p & r = q \\ q < p & r = \frac{q}{p} \end{cases} \tag{20}$$

p can be determined as the ratio of non-tail observations in total observations, or as 1 minus the share of tail observations in total observations. It is necessary to know the correct number and the corresponding positions of observations below a specific transition value $x_0 > x_{min}$ and the cumulative density function at value x_0, gives the share of observations below x_0 in the tail. From the latter, we can derive the number of total tail observations. Then p corresponds to the share of non-tail observations in total observations. Let $n(x_i)$ denote the number of observations with value x_i, i.e. survey weights.

$$\sum_{x_i > x_{min}} = \frac{1}{F(w_0)} * \sum_{x_i \in [x_{min}, x_0]}$$
$$p = \frac{\sum_{x_i < x_{min}} n(x_i)}{\sum_{x_i < x_{min}} n(x_i) + \frac{1}{F(w_0)} * \sum_{x_i \in [x_{min}, x_0]}} \tag{21}$$

Then, the top share as a ratio of conditional means depends on distribution-specific conditional means and quantile functions.

A.2 Parametrical Solution for the Pareto Tail

The quantile distribution for the Pareto function is defined as $Q(X,r) = F^{-1}(X,r) = x_{min}(1-r)^{-1/\alpha}$. The conditional expected value is $\bar{X} \mid X \geq Q(X,r) = \frac{\alpha}{1-\alpha} Q(X,r)$. Thus, the top share for a percentile q in a mixed distribution with a Pareto tail is given by:

$$TS(X,q,p) = \begin{cases} \frac{(1-q)\frac{\alpha}{\alpha-1}(1-\frac{q-p}{1-p})^{-1/\alpha} x_{min}}{p[\bar{X}|X<x_{min}] + (1-p)\frac{\alpha}{\alpha-1} x_{min}} & q > p \\ \frac{(1-p)\frac{\alpha}{\alpha-1} x_{min}}{p[\bar{X}|Y<x_{min}] + (1-p)\frac{\alpha}{\alpha-1} x_{min}} & p = q \\ \frac{(p-q)[\bar{X}|X \in [Q^{emp}, x_{min}]] + (1-p)\frac{\alpha}{\alpha-1} x_{min}}{p[\bar{X}|X<x_{min}] + (1-p)\frac{\alpha}{\alpha-1} x_{min}} & q < p \end{cases} \tag{22}$$

A.3 Parametrical Solution for the Generalized Pareto Tail

For the Generalized Pareto, the conditional mean $E[\bar{X} \mid X > Q(X,r)] = Q(X,r) + \frac{\sigma + \xi(Q(X,r)-\mu)}{1-\xi}$ follows from the empirical excess function $E[X-u \mid X > u] = \frac{\sigma+\xi(u-\mu)}{1-\xi}$ (Langousis et al., 2016, p. 2664). The quantile function follows from the inverse cumulative density function $Q(X,r) = F^{-1}(r) = \frac{(1-r)^{-\xi} \times (\sigma + \mu\xi(1-r)^{\xi} - \sigma \times (1-r))^{\xi}}{\xi}$. The mean tail observation is given by the special case $Q(X,r) = \mu$: $E[X \mid X > \mu] = \mu + \frac{\sigma}{1-\xi}$. The top share for a percentile q in a mixed distribution with a generalized Pareto tail is given by:

$$TS(X,q,p) = \begin{cases} \frac{q\left[Q(X,r)\frac{\sigma+\xi(Q(X,r)-\mu)}{1-\xi}\right]}{p[\bar{X}|X<\mu] + (1-p)\left[\mu + \frac{\sigma}{1-\xi}\right]} & q > p \\ \frac{p[\bar{X}|X<\mu]}{p[\bar{X}|X<\mu] + (1-p)\left[\mu + \frac{\sigma}{1-\xi}\right]} & q = p \\ \frac{(p-q)[\bar{X}|X \geq Q^{emp}(X,r) + p\left[\mu + \frac{\sigma}{1-\xi}\right]]}{p[\bar{X}|X<\mu] + (1-p)\left[\mu + \frac{\sigma}{1-\xi}\right]} & q < p \end{cases} \tag{23}$$
$$Q(X,r) = \frac{(1-r)^{-\xi} \times (\sigma + \mu\xi(1-r)^{\xi} - \sigma \times (1-r))^{\xi}}{\xi} \tag{24}$$

References

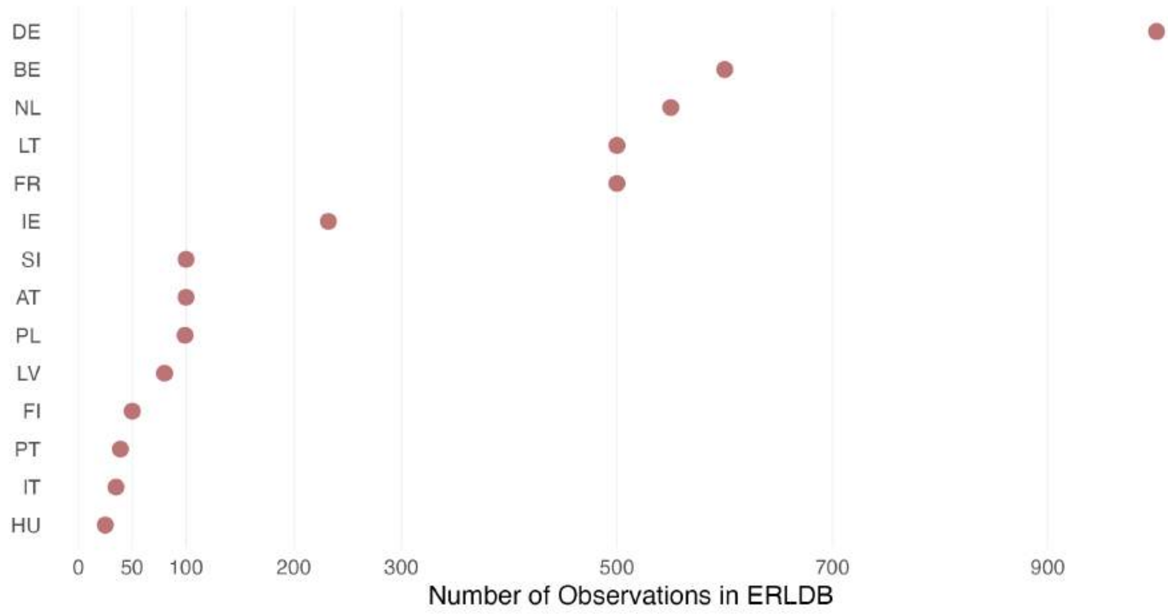
Charpentier, A. and Flachaire, E. (2022): Pareto models for top incomes and wealth. In: The Journal of Economic Inequality. Vol. 20 (1), 1-25.

Langousis, A.; Mamalakis, A.; Puliga, M. and Deidda, R. (2016): Threshold detection for the Generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. In: Water Resources Research. Vol. 52 (4), 2659–2681.

B Supporting Material: Data and Main Results

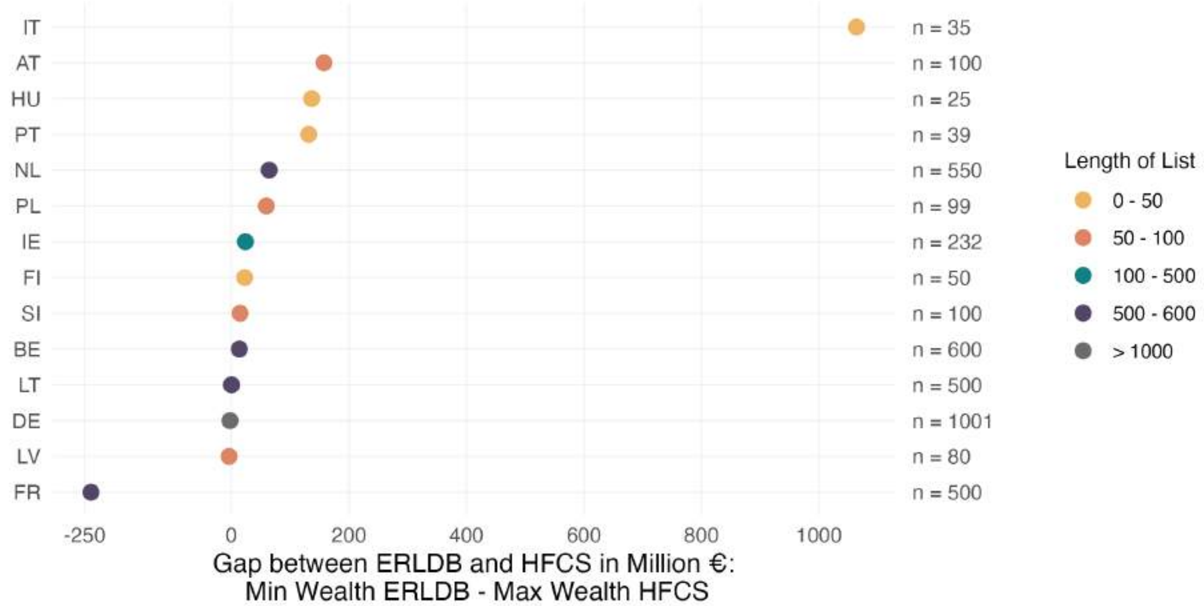
Figure B.1.	**Length of Rich List by Country**	**62**
Figure B.2.	**Micro-Micro Gap by Length of the Rich List**	**63**
Figure B.3.	w_0 **and the Effective Oversampling Rate**	**64**
Table B.1.	**Summary Table of HFCS 2017 and ERLDB data.**	**65**
Table B.2.	**Main Results: Parameter Estimates**	**66**
Table B.3.	**Aggregate Wealth Compared to Macroeconomic Aggregates**	**66**

Figure B.1: Length of Rich List by Country



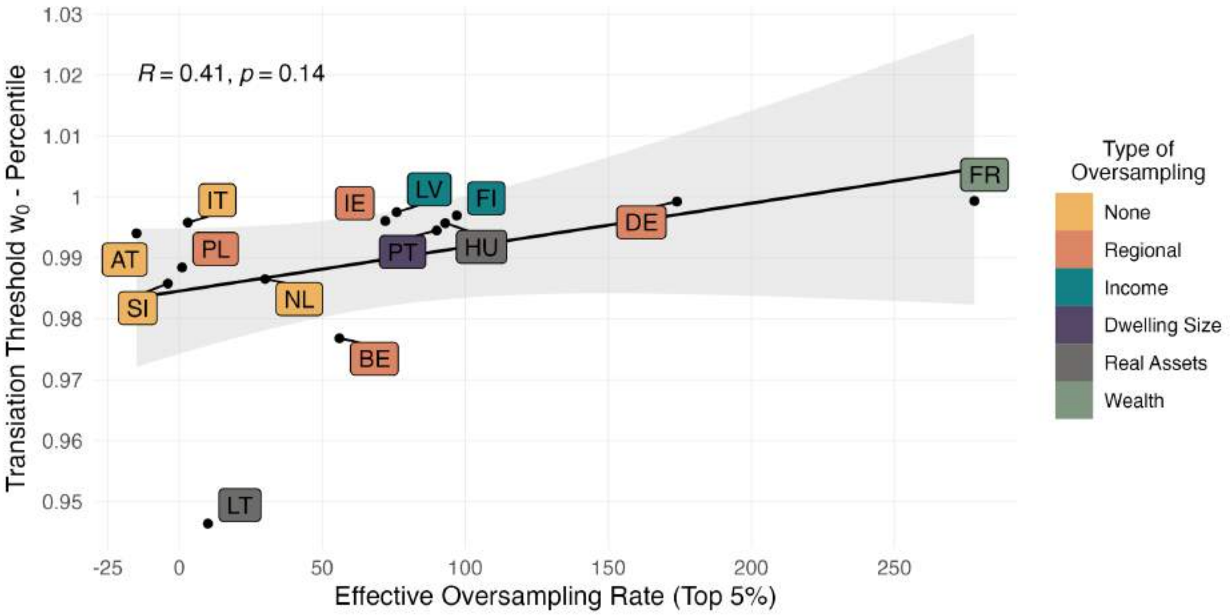
Note: This figure shows the number of entries of each rich list.

Figure B.2: Micro-Micro Gap by Length of the Rich List



Note: This figure shows the gap between the maximum wealth according th HFCS and the lowest wealth recorded in the country-specific rich list by country and length of the list.

Figure B.3: w_0 and the Effective Oversampling Rate



Note: This figure shows a positive correlation between the oversampling rate of the top 5% in HFCS and the transition parameter w_0. Higher oversampling of rich households in the survey thus corresponds that is location at higher percentiles of the wealth distribution.

Table B.1: Summary Table of HFCS 2017 and ERLDB

	Household Finance and Consumption Survey (HFCS)							European Rich List Database (ERLDB)				
				Net wealth (€)		Oversampling					Wealth (m€)	
	Field Period	Ref. Year	Sample Size	Mean	Median	Type	Effective Rate Top 5%	Source	Year	Sample Size	Min.	Max.
AT	11/2016–07/2017	2017	3,072	250,300	82,700	none	-15%	Trend	2017	100	200	35,400
BE	01/2017–09/2017	2017	2,329	366,200	212,500	[regional] areas with higher number of households and larger dispersion of income	56%	De Rijkste Belgen	2018	600	25	17,295
DE	03/2017–10/2017	2017	4,942	232,800	70,800	[regional] wealthy street sections in cities, municipalities with a high share of taxpayers with a certain income	174%	Manager Magazin	2017	1001	90	33,000
FI	01/2017–06/2017*	2016	1,021	206,600	107,200	[income] register data	97%	Arvopaperi	2016	50	31	1,490
FR	09/2017–01/2018	2017	13,685	242,000	117,600	[wealth] register data	278%	Challenges	2017	500	130	46,900
HU	10/2017–12/2017	2017	5,968	71,800	35,900	[dwellings]	93%	Napi	2019	25	148	1,107
IE	04/2018–01/2019	2018	4,793	367,800	185,000	[regional] areas with high wealth index based on homeownership rates and local property tax bands	72%	Sunday Independent	2018	232	50	15,600
IT	01/2017–09/2017*	2016	742	214,300	132,300	none	3%	Forbes Italia	2019	35	1,072	20,018
LV	09/2017–11/2017	2017	1,249	43,000	20,500	[income] register data	76%	Dienas Bizness	2017	80	9	172
LT	12/2017–05/2018*	2016	1,664	84,300	45,900	[wealth] real assets from register data	10%	Alfa	2019	500	2.1	1,400
NL	05/2017–07/2017	2017	2,556	186,000	67,400	none	30%	Quote	2018	550	80	12,800
PL	09/2016–11/2016	2016	5,858	95,500	60,500	[income, property] property size and register data on income	0%	wprost	2016	100	63.4	3,641
PT	05/2017–09/2017	2017	5,924	162,300	74,800	[dwellings] size of dwelling	90%	Forbes	2018	39	155	4,502
SI	04/2017–10/2017	2017	2,014	144,300	91,600	none	-4%	Finance Manager	2018	100	24.2	689

*) Assets and liabilities are reported as of 31 December 2016 and not at the time of the interview. Based on European Central Bank (2020).

Table B.2: Main Results: w_{min}, w_0 and (Generalized) Pareto Distribution Parameters

	w_{min}		w_0		Alpha		Scale	
	Absolute	*Relative*	*Absolute*	*Relative*	*Pareto*	*GPareto*	*Pareto*	*GPareto*
AT	231,600	0.694	2,945,760	0.994	1.315	1.449	159,855	175,183
BE	270,000	0.579	1,979,800	0.977	1.609	1.658	162,817	207,552
DE	314,600	0.786	8,035,880	0.999	1.400	1.628	193,234	218,406
FI	330,000	0.813	2,827,000	0.997	1.886	2.087	158,118	201,258
FR	403,000	0.836	9,039,400	0.999	1.730	1.582	254,706	224,855
HU	69,600	0.734	1,083,360	0.996	1.508	1.642	42,386	54,013
IE	765,600	0.872	4,428,420	0.996	1.473	1.582	483,826	499,495
IT	268,000	0.762	2,208,800	0.996	1.644	2.395	111,901	201,717
LT	36,400	0.392	270,560	0.946	1.377	1.540	23,633	37,053
LV	29,800	0.605	931,580	0.997	1.412	1.655	18,002	27,182
NL	257,800	0.790	1,519,720	0.987	1.327	1.503	171,523	173,443
PL	96,200	0.677	580,640	0.988	1.634	2.084	46,162	63,121
PT	160,000	0.750	2,033,680	0.994	1.465	1.636	97,773	133,952
SI	147,200	0.704	890,140	0.986	1.566	1.827	80,590	100,640

Note: This table is based on all five implicates of HFCS 2017 data.

Table B.3: Aggregate Wealth Compared to Macroeconomic Aggregates.

	Nat. accounts	HFCS		Pareto		GPareto	
		Absolute	*Relative*	*Absolute*	*Relative*	*Absolute*	*Relative*
AT	1,498,993	984,564	65.7	1,293,681	86.3	1,107,142	73.9
BE	2,516,688	1,788,913	71.1	1,806,266	71.8	1,844,012	73.3
DE	12,371,259	9,394,146	75.9	11,633,664	94	9,922,732	80.2
FI	600,821	553,060	92.1	563,735	93.8	564,383	93.9
FR	11,375,520	7,096,665	62.4	7,210,706	63.4	7,544,517	66.3
HU	356,522	287,688	80.7	297,767	83.5	295,537	82.9
IT	9,516,027	5,468,243	57.5	6,149,729	64.6	5,548,933	58.3
LT	79,287	108,435	136.8	115,934	146.2	120,791	152.3
LV	97,567	36,018	36.9	40,820	41.8	37,801	38.7
NL	3,346,519	1,449,603	43.3	2,048,527	61.2	1,636,048	48.9
PL	632,270	1,277,826	202.1	1,440,439	227.8	1,310,758	207.3
PT	789,194	668,212	84.7	695,641	88.1	668,825	84.7
SI	126,593	119,010	94	132,397	104.6	122,238	96.6

Note: Absolute values are in €million, relative values are in relation to national accounts. This table is based on all five implicates of HFCS 2017 data.

C CCDF Plots by Country and Implicate

Figure C.1: CCDF and Parameter Estimates: AT

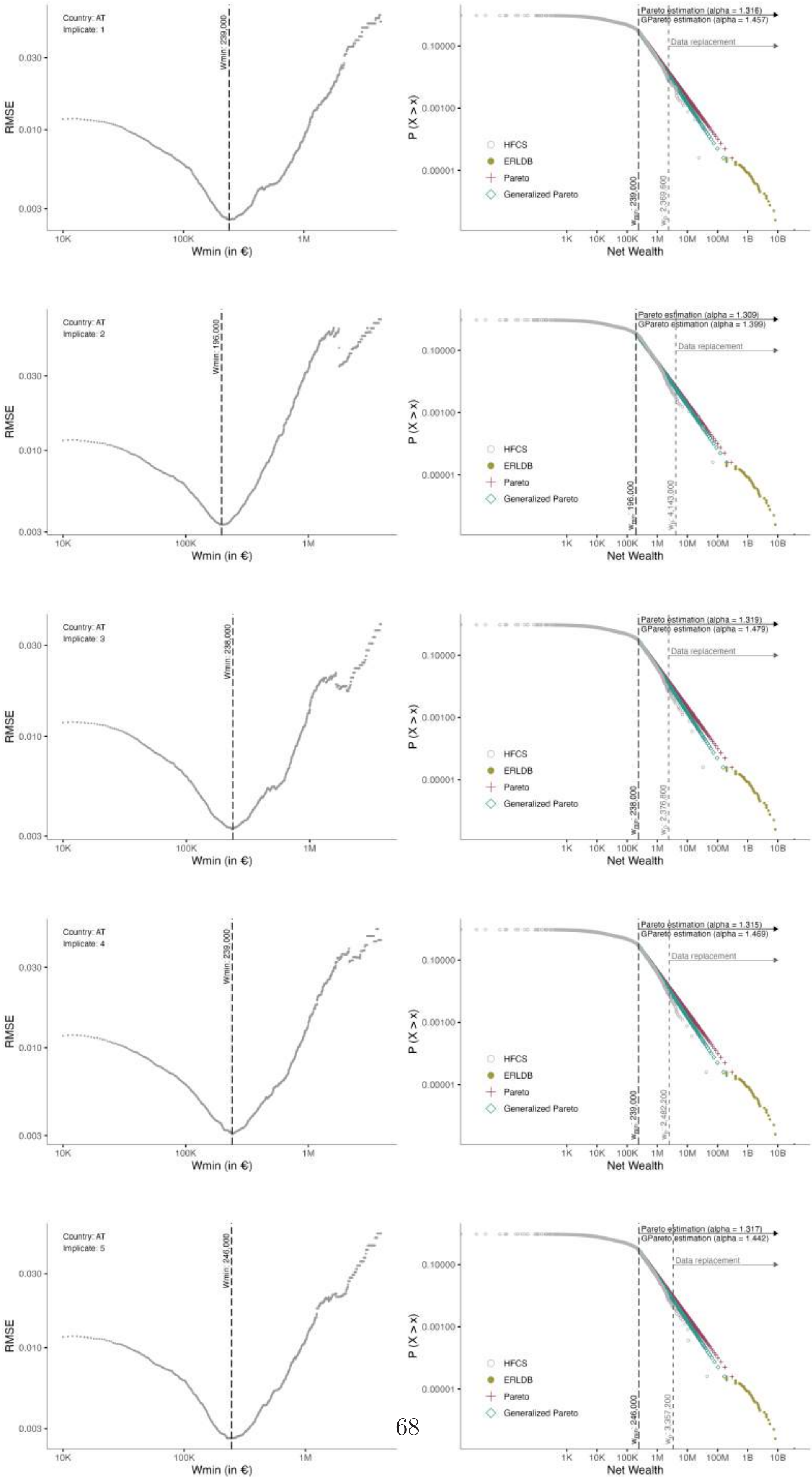


Figure C.2: CCDF and Parameter Estimates: BE

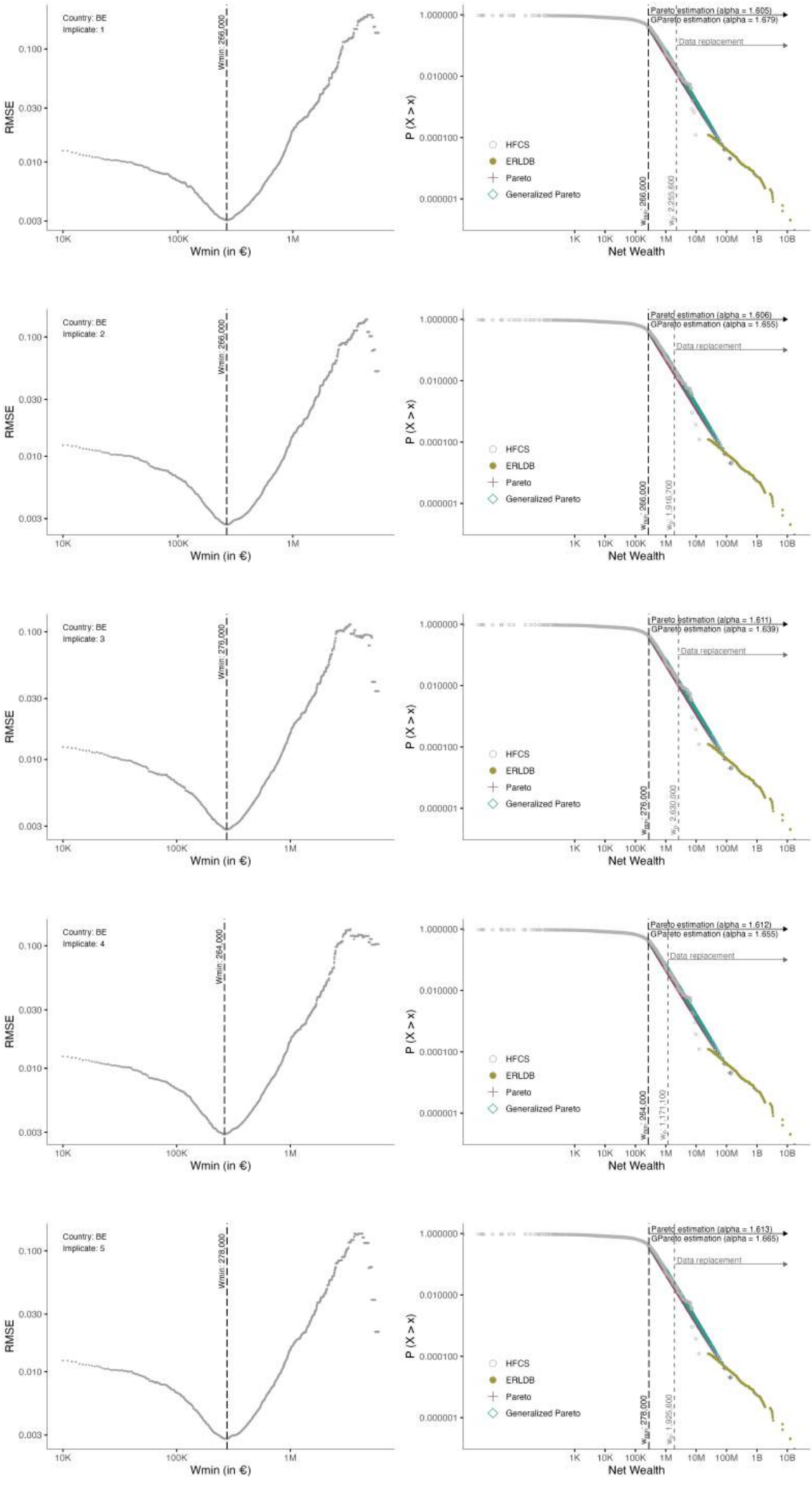


Figure C.3: CCDF and Parameter Estimates: DE

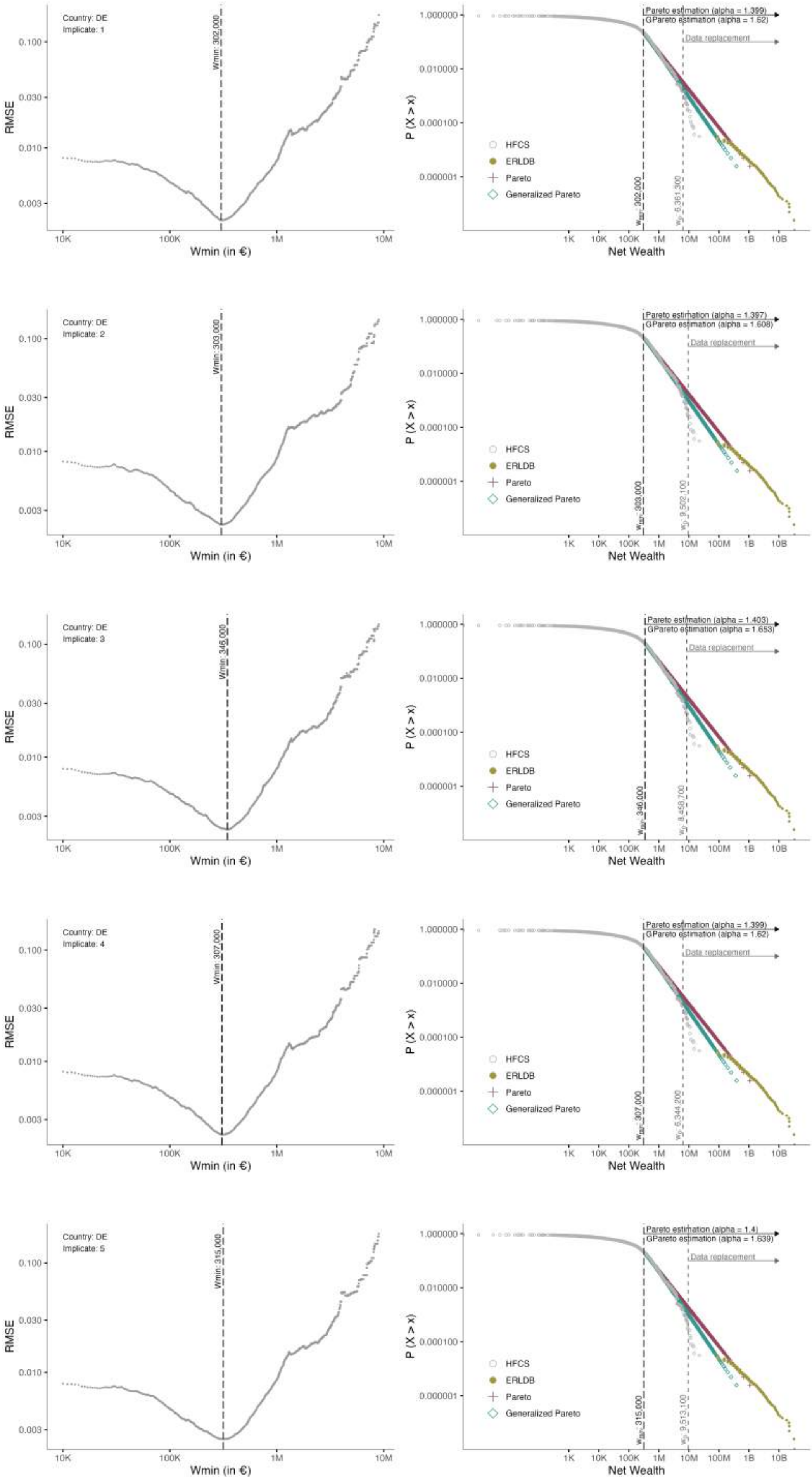


Figure C.4: CCDF and Parameter Estimates: FI

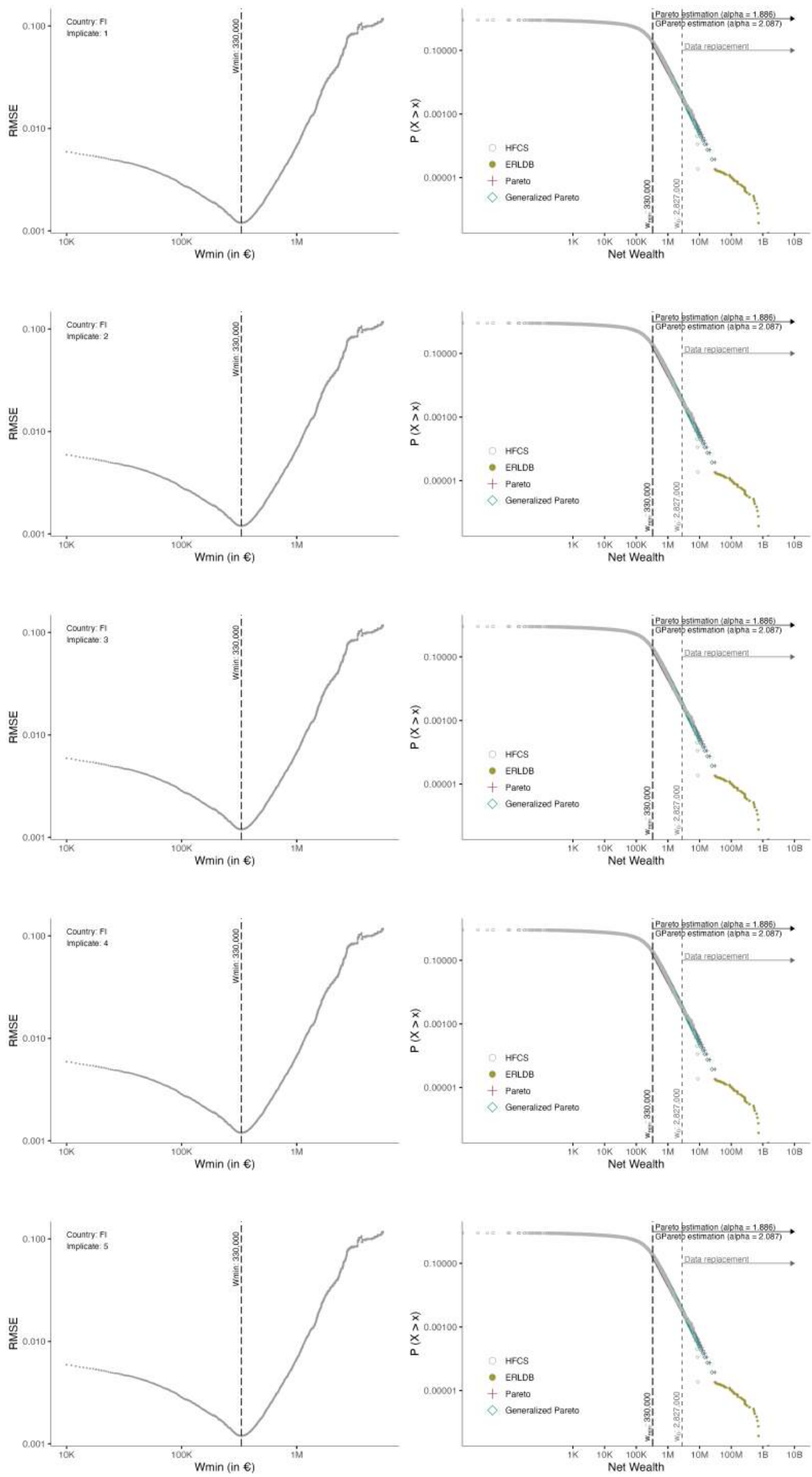


Figure C.5: CCDF and Parameter Estimates: FR

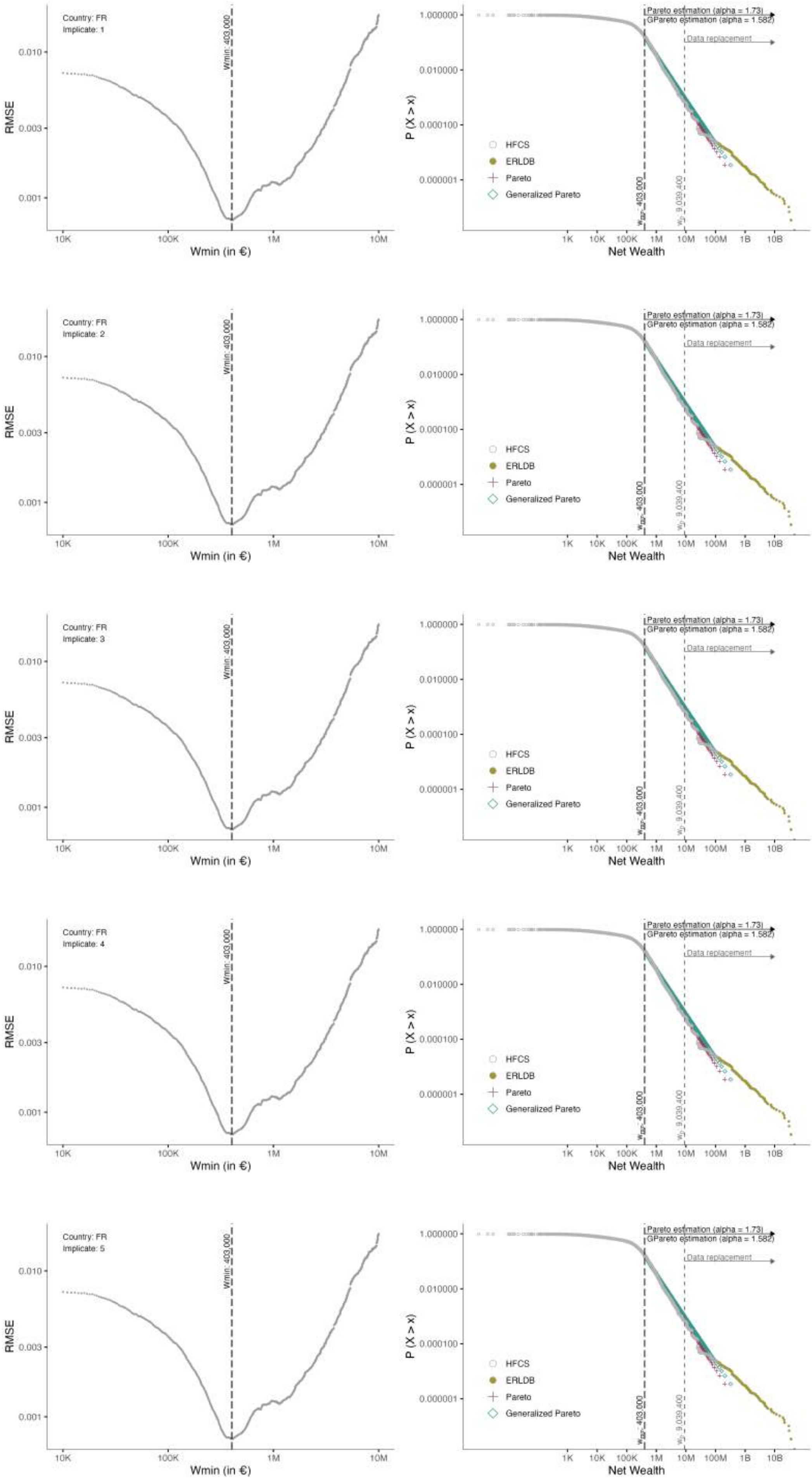


Figure C.6: CCDF and Parameter Estimates: HU

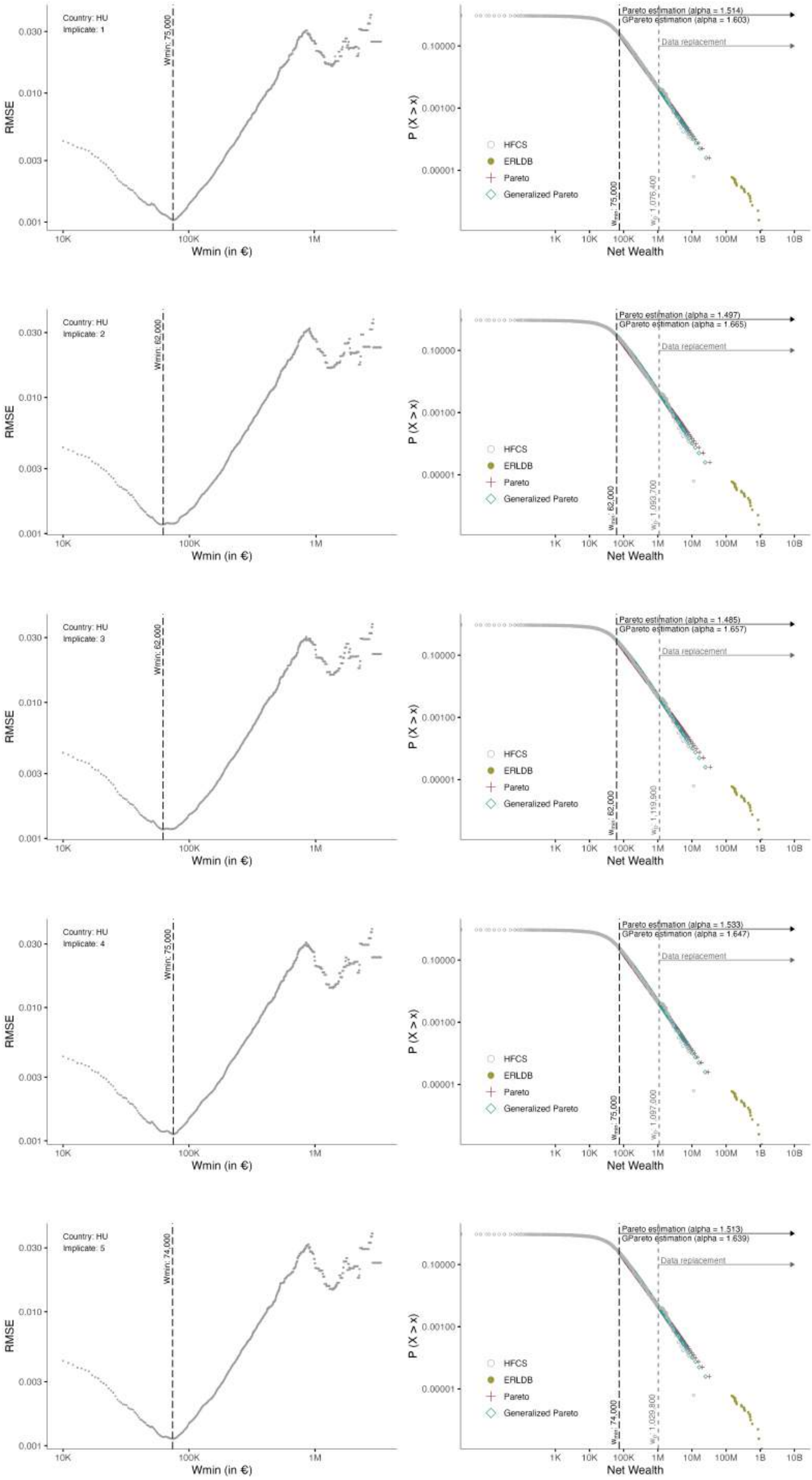


Figure C.7: CCDF and Parameter Estimates: IE

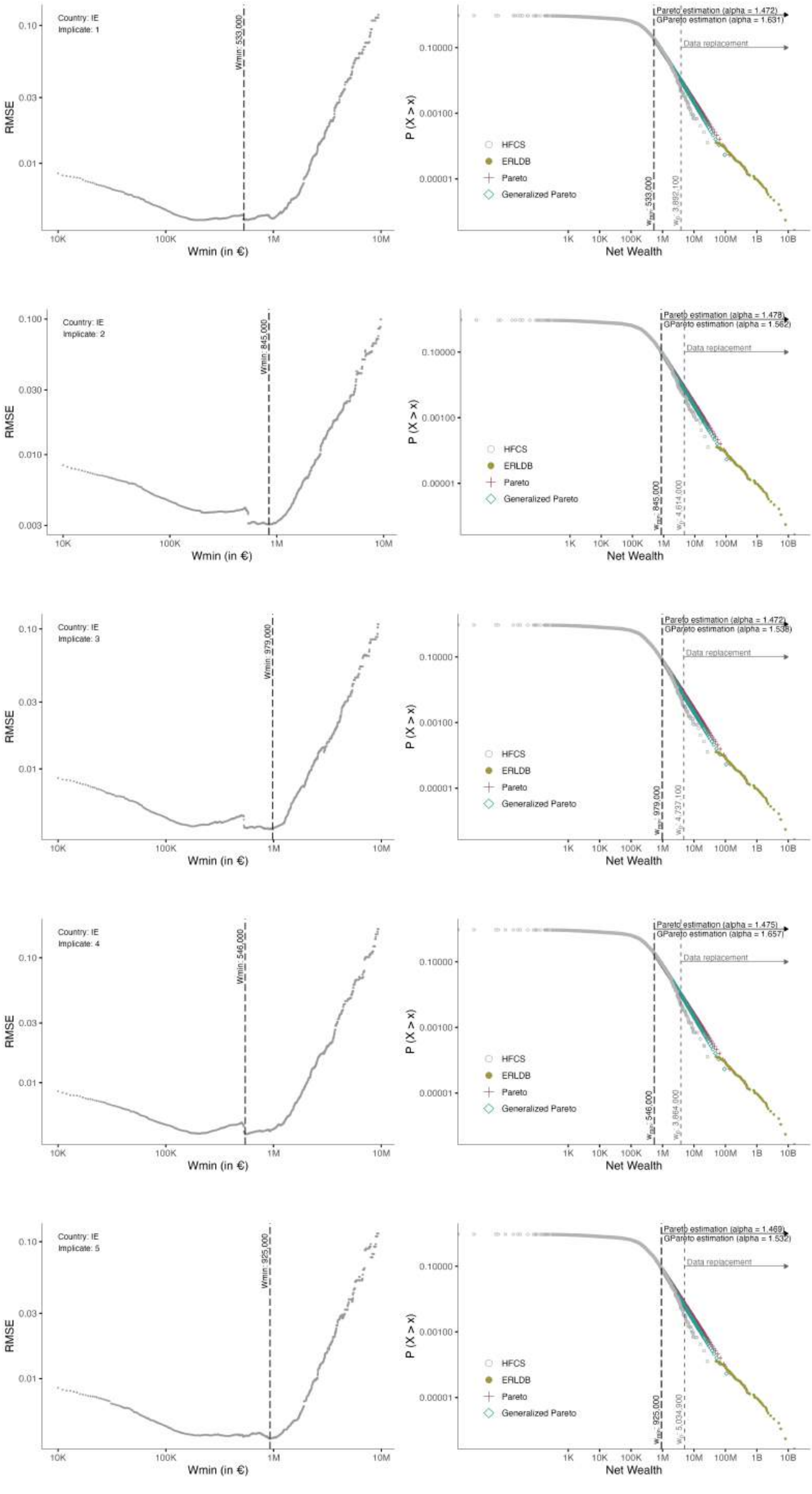


Figure C.8: CCDF and Parameter Estimates: IT

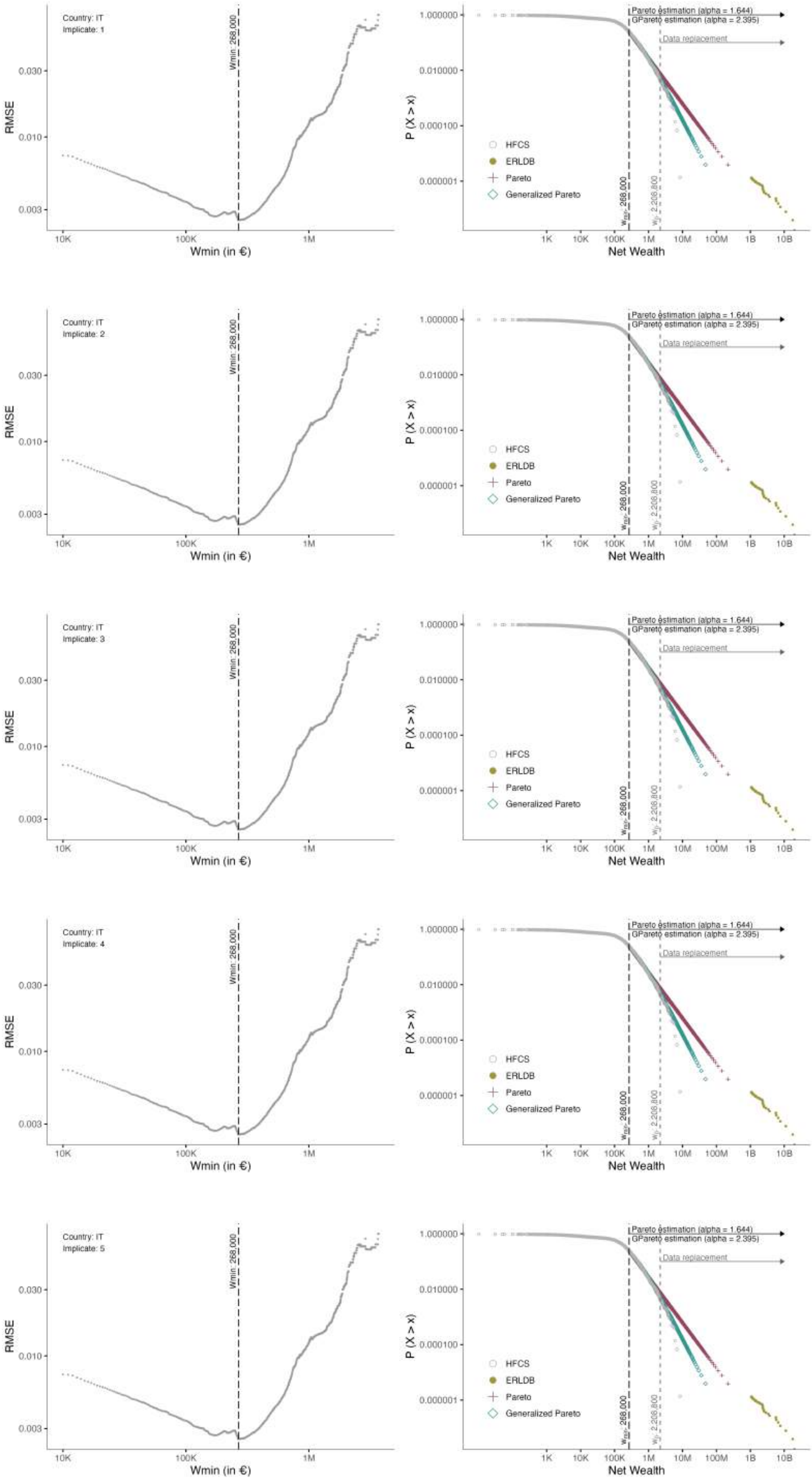


Figure C.9: CCDF and Parameter Estimates: LT

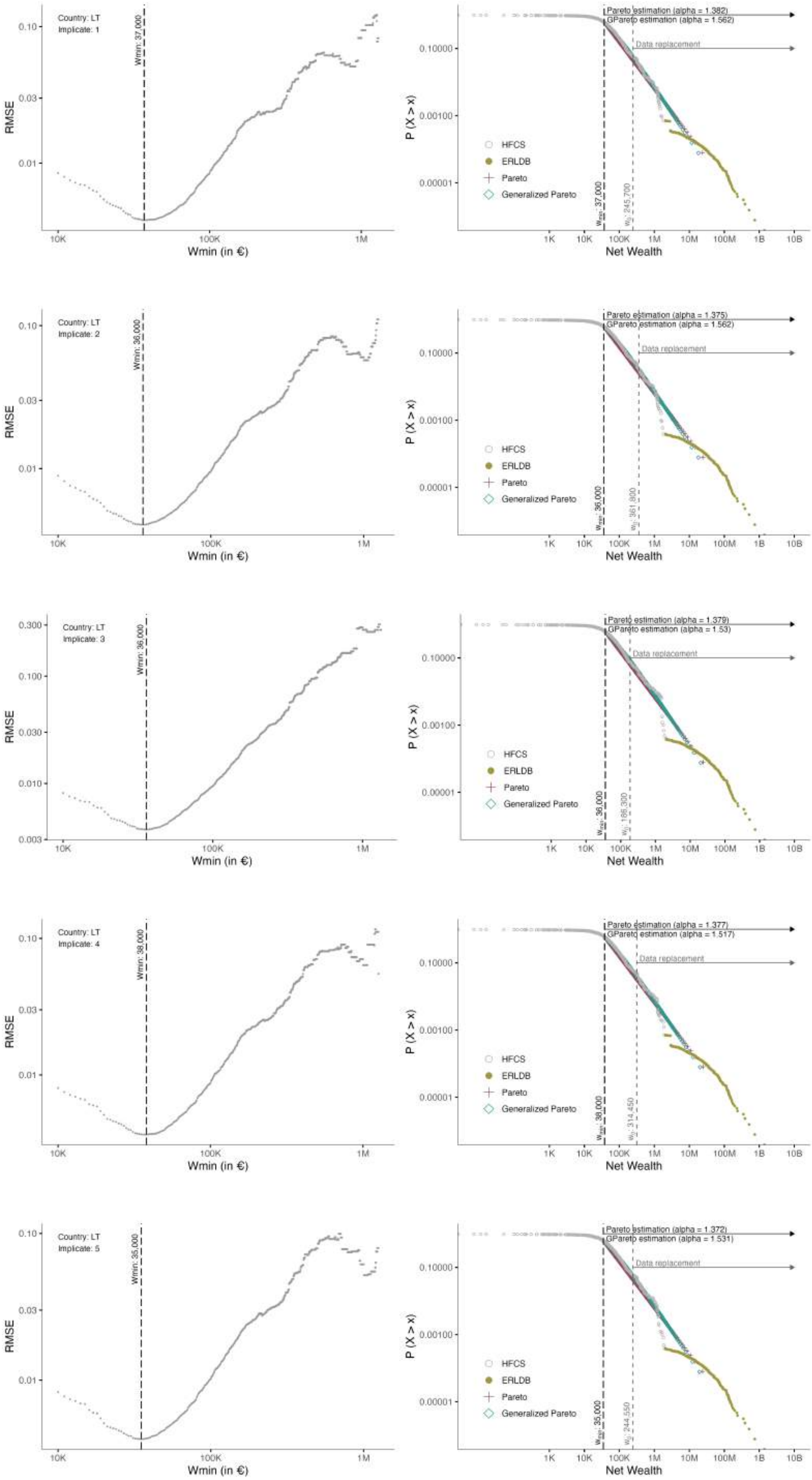


Figure C.10: CCDF and Parameter Estimates: LV

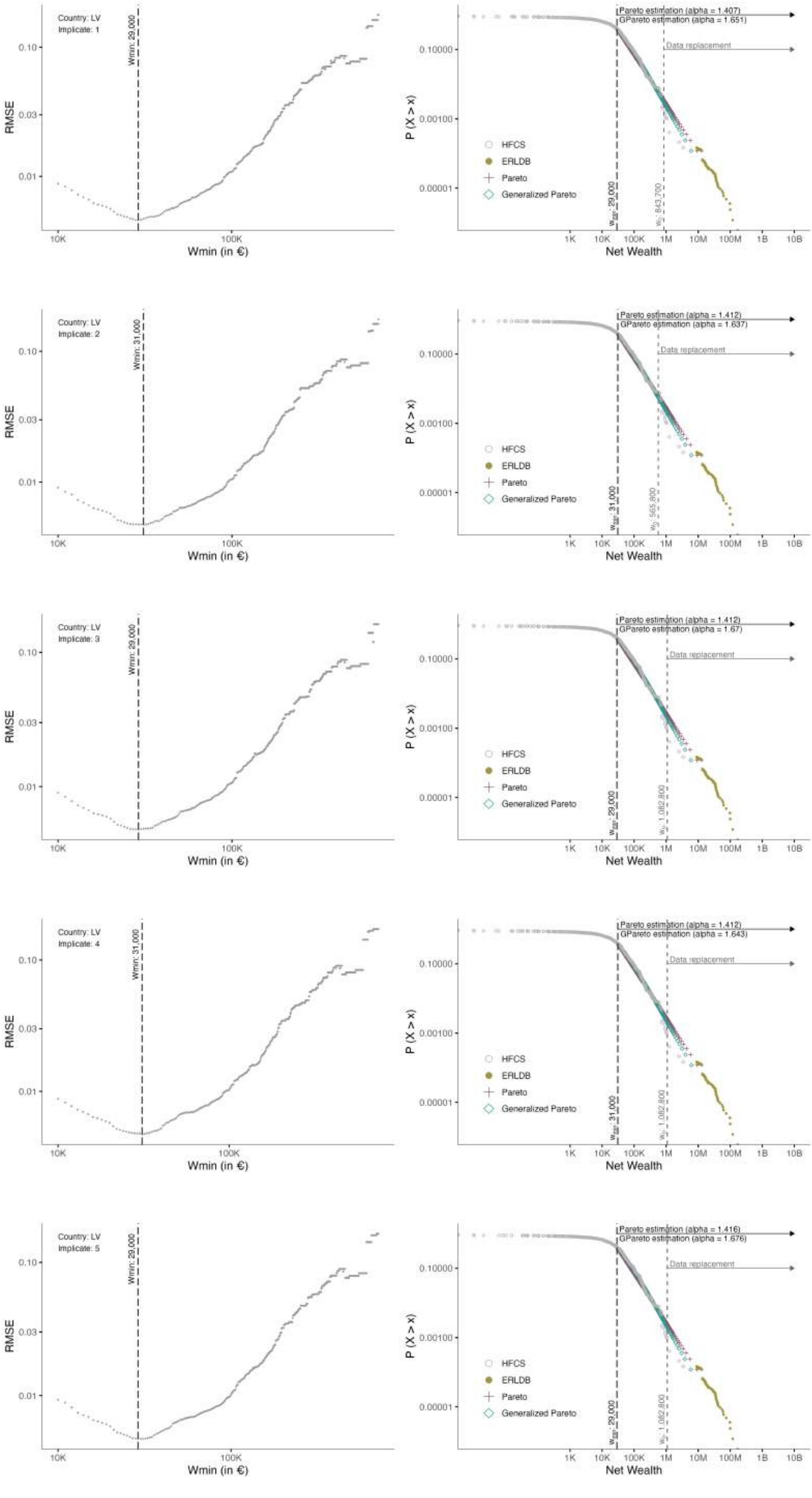


Figure C.11: CCDF and Parameter Estimates: NL

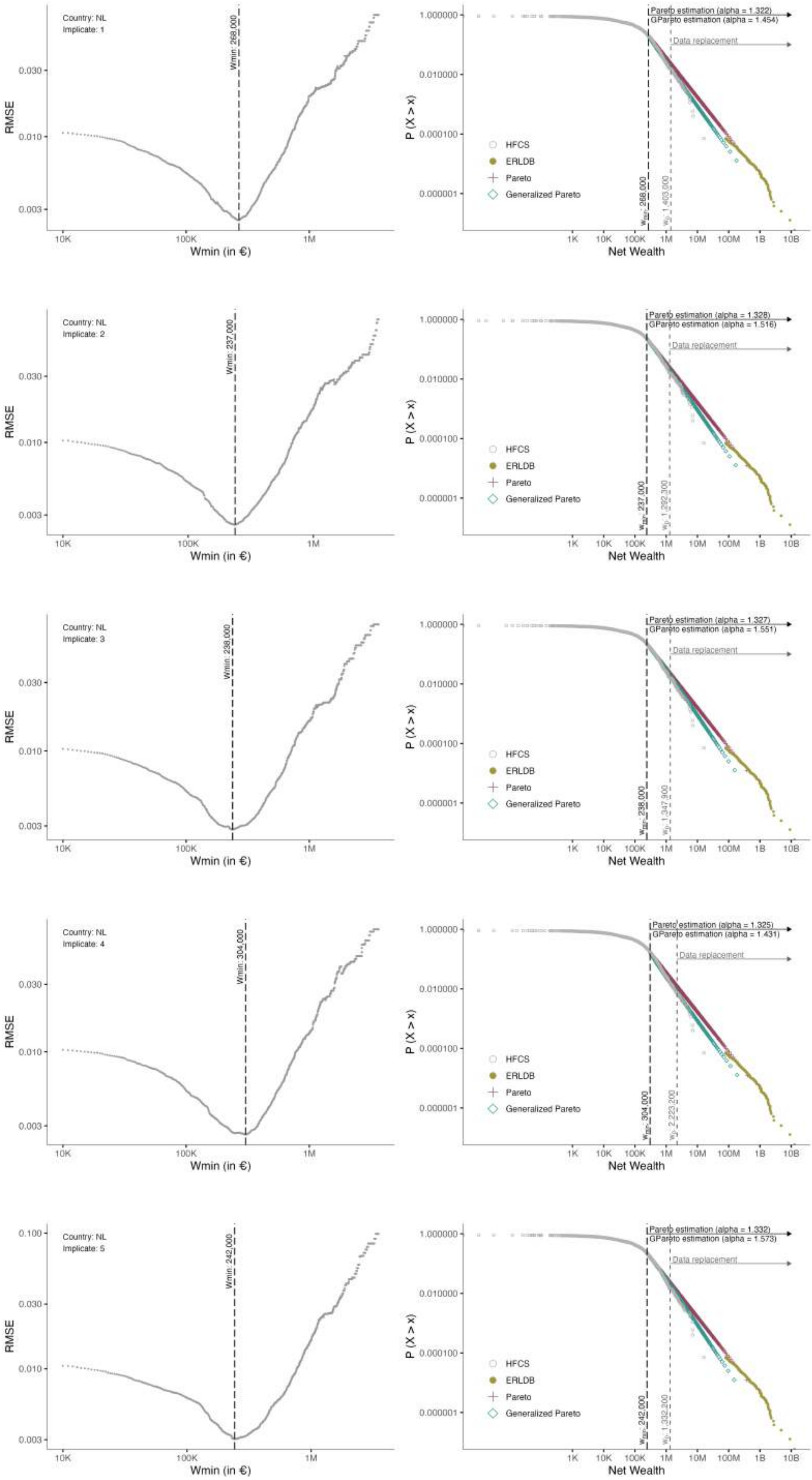


Figure C.12: CCDF and Parameter Estimates: PL

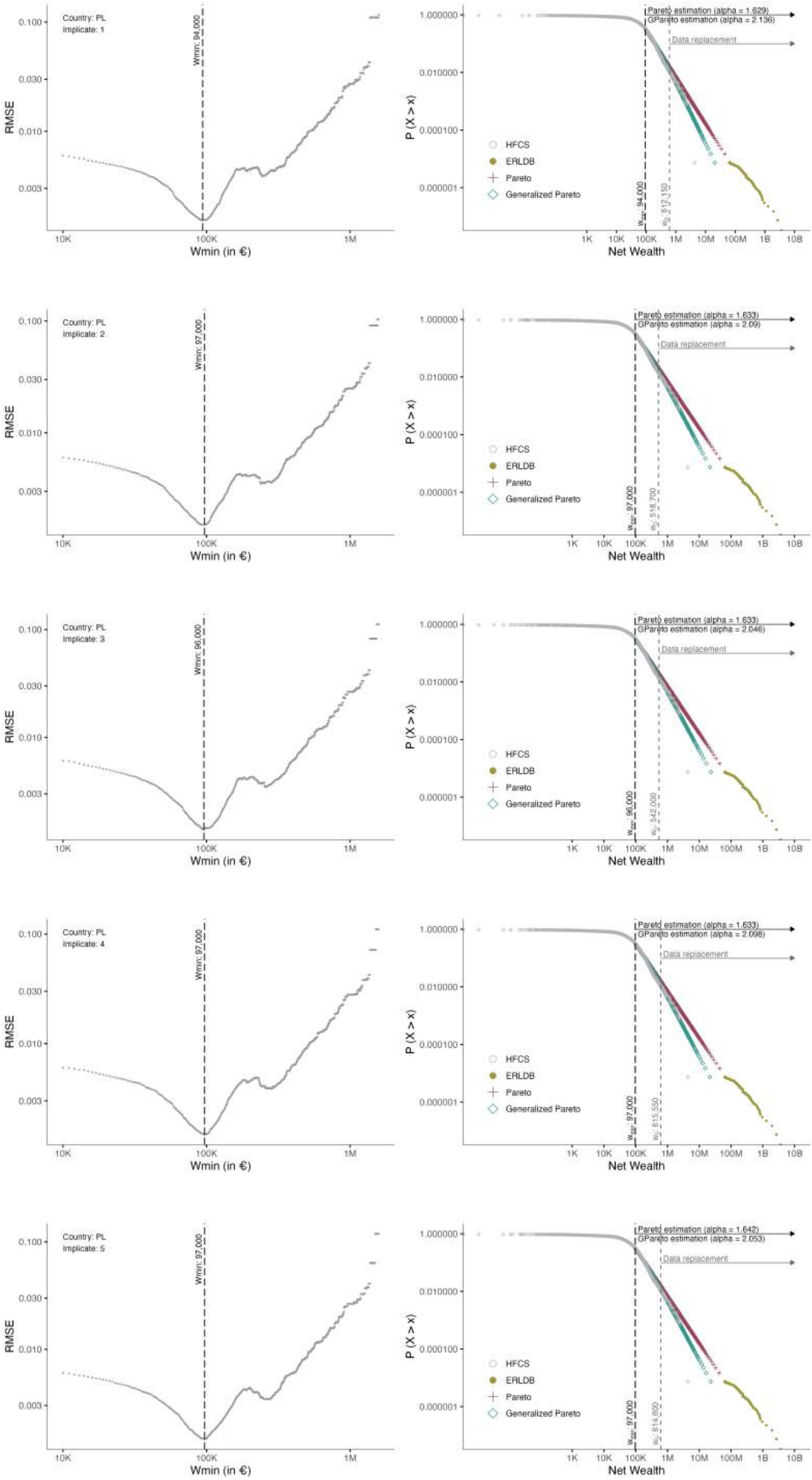


Figure C.13: CCDF and Parameter Estimates: PT

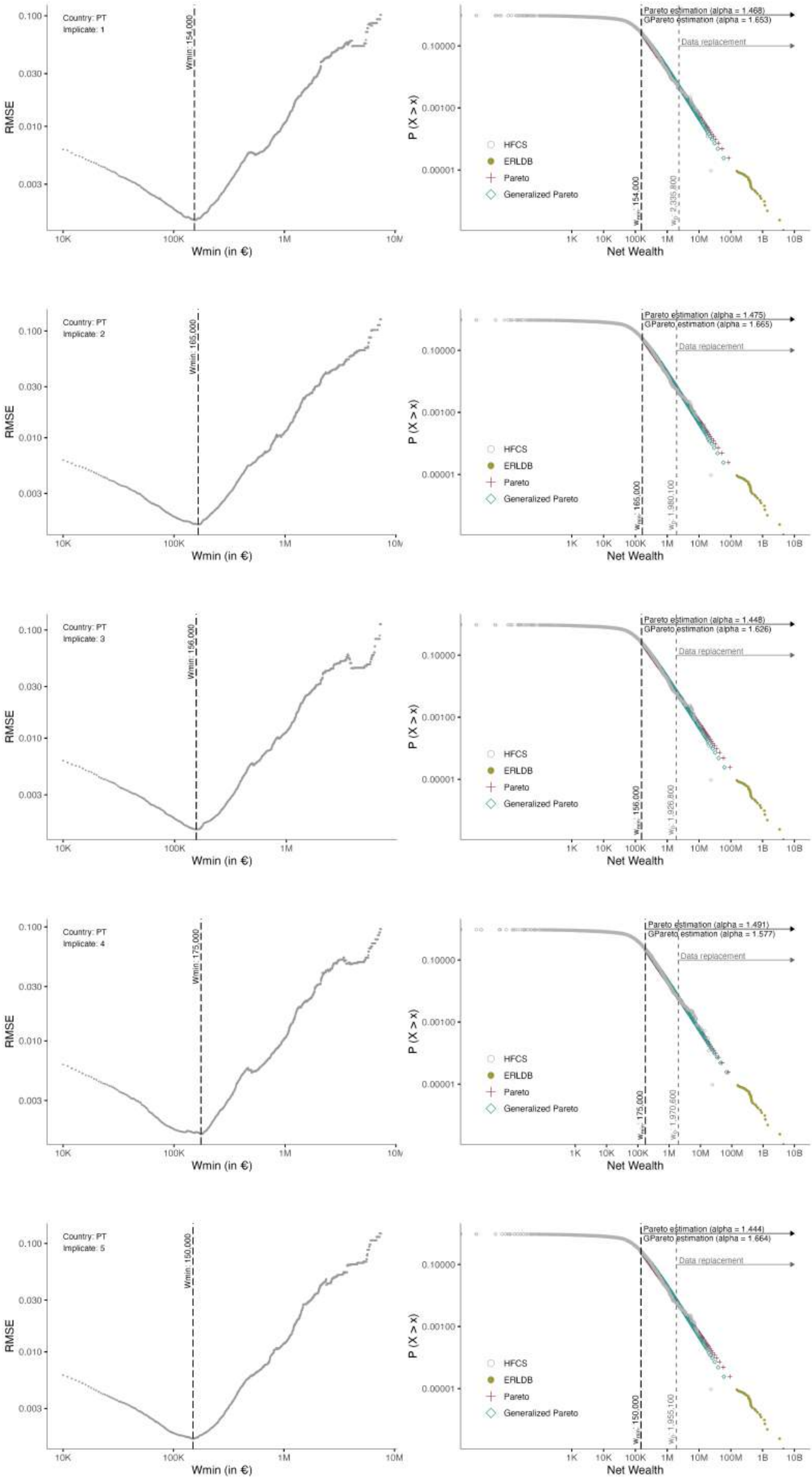
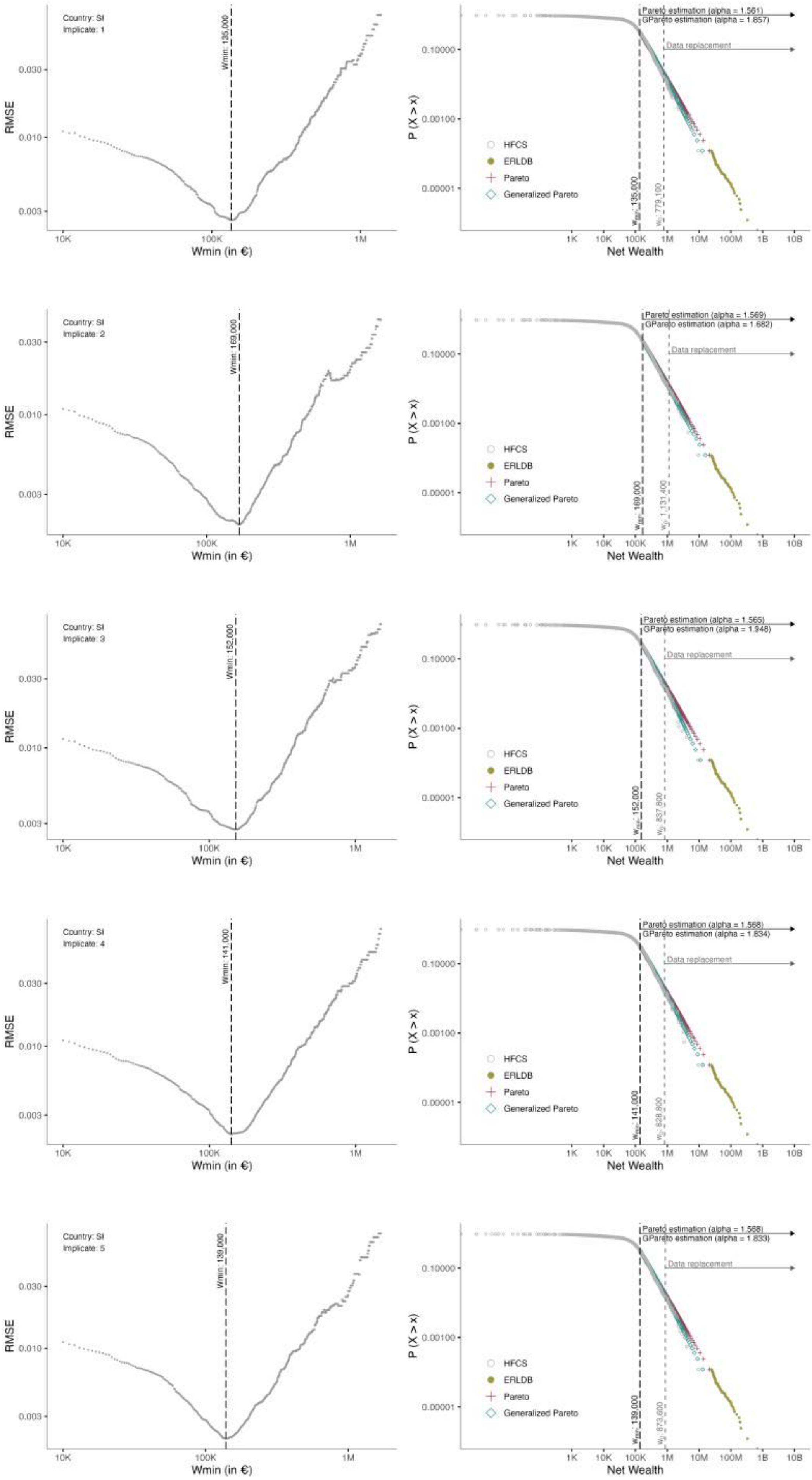


Figure C.14: CCDF and Parameter Estimates: SI



D Transition Threshold Parameter Determination by Country

Figure D.1: Determination of Transition Threshold Parameter w_0 - AT

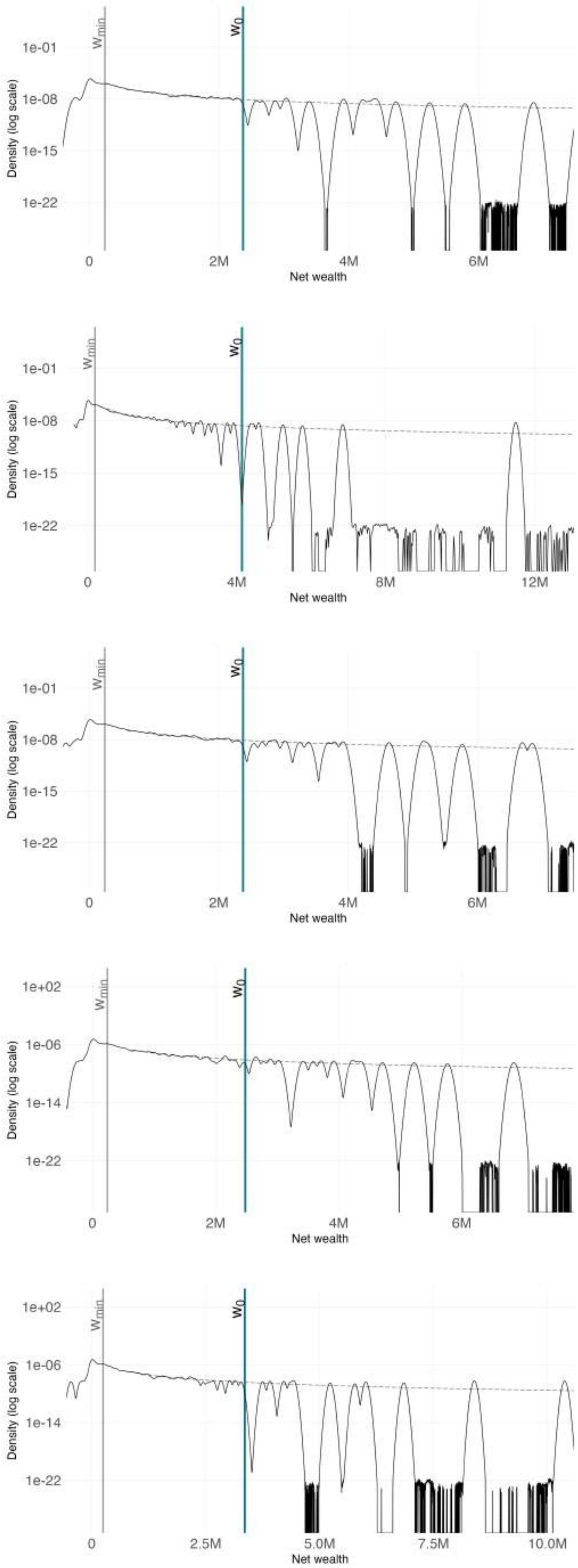


Figure D.2: Determination of Transition Threshold Parameter w_0 - BE

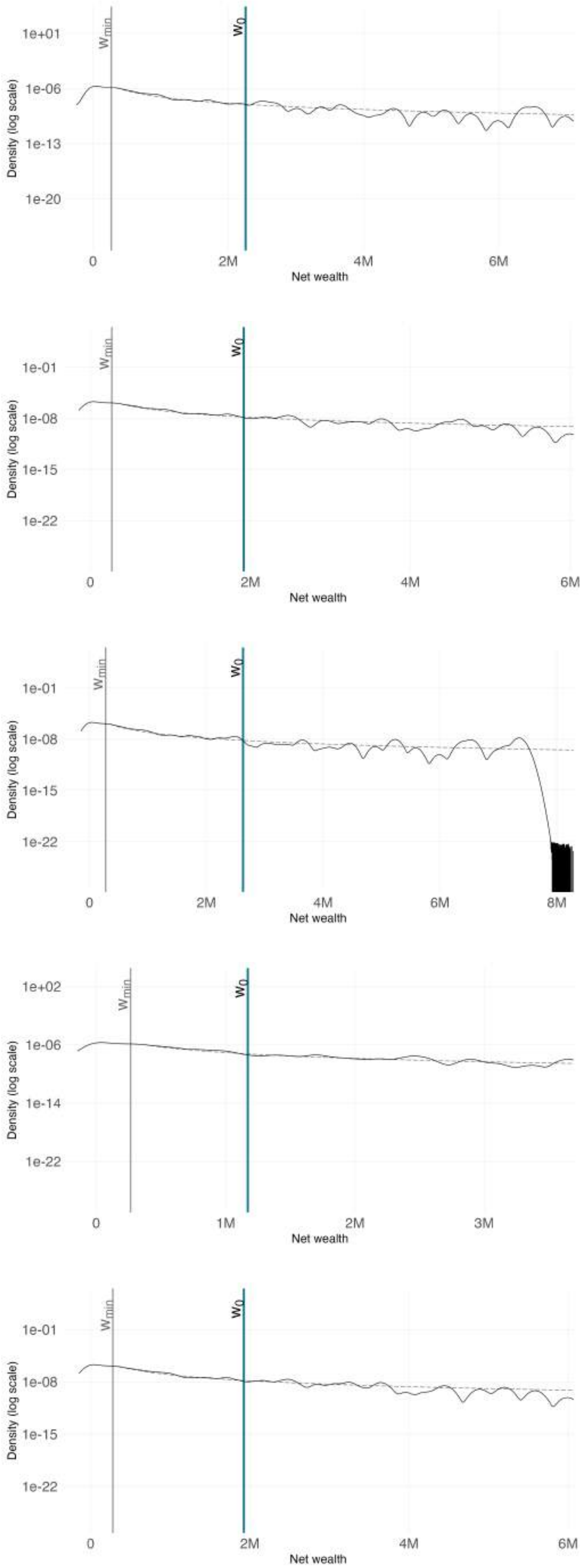


Figure D.3: Determination of Transition Threshold Parameter w_0 - DE

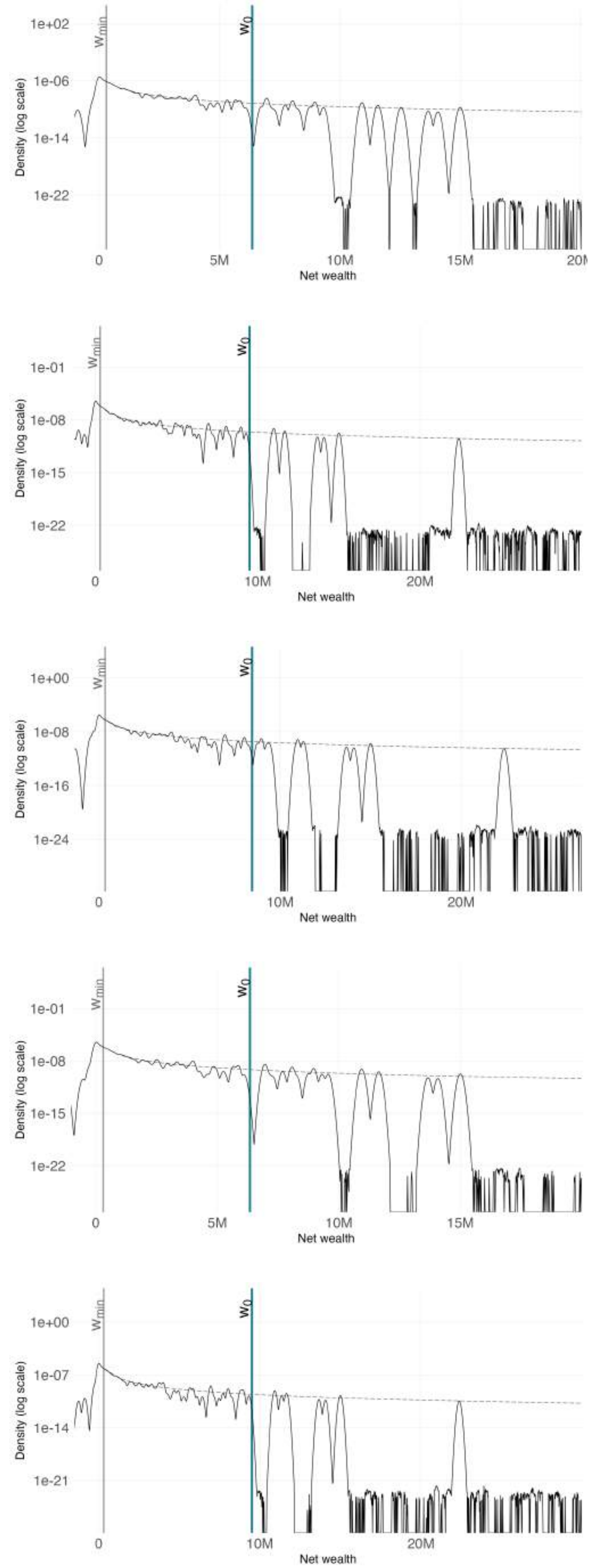


Figure D.4: Determination of Transition Threshold Parameter w_0 - FI

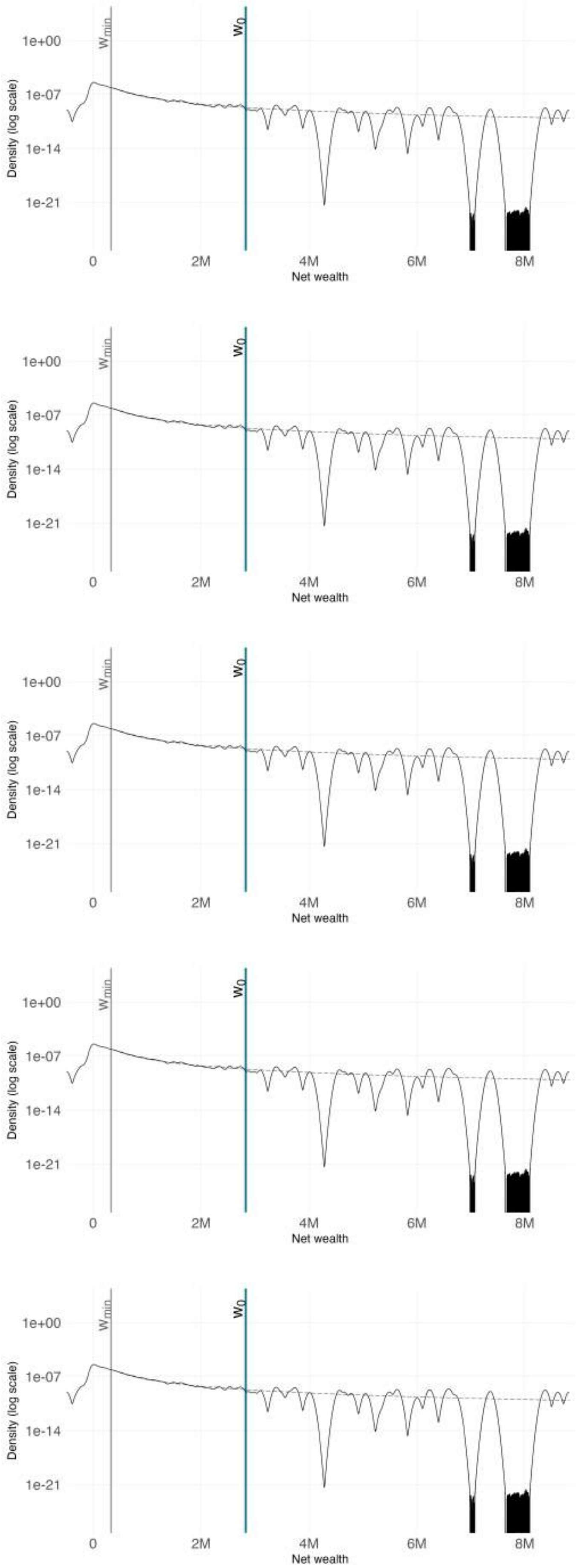


Figure D.5: Determination of Transition Threshold Parameter w_0 - FR

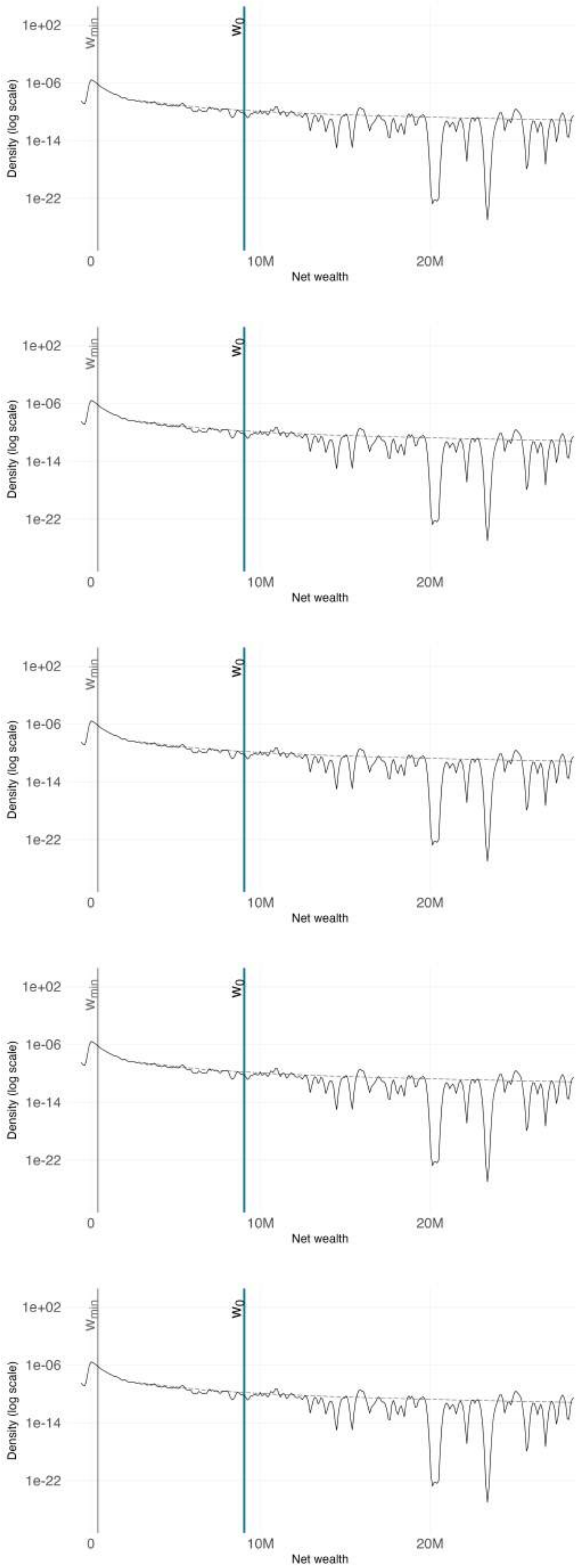


Figure D.6: Determination of Transition Threshold Parameter w_0 - HU

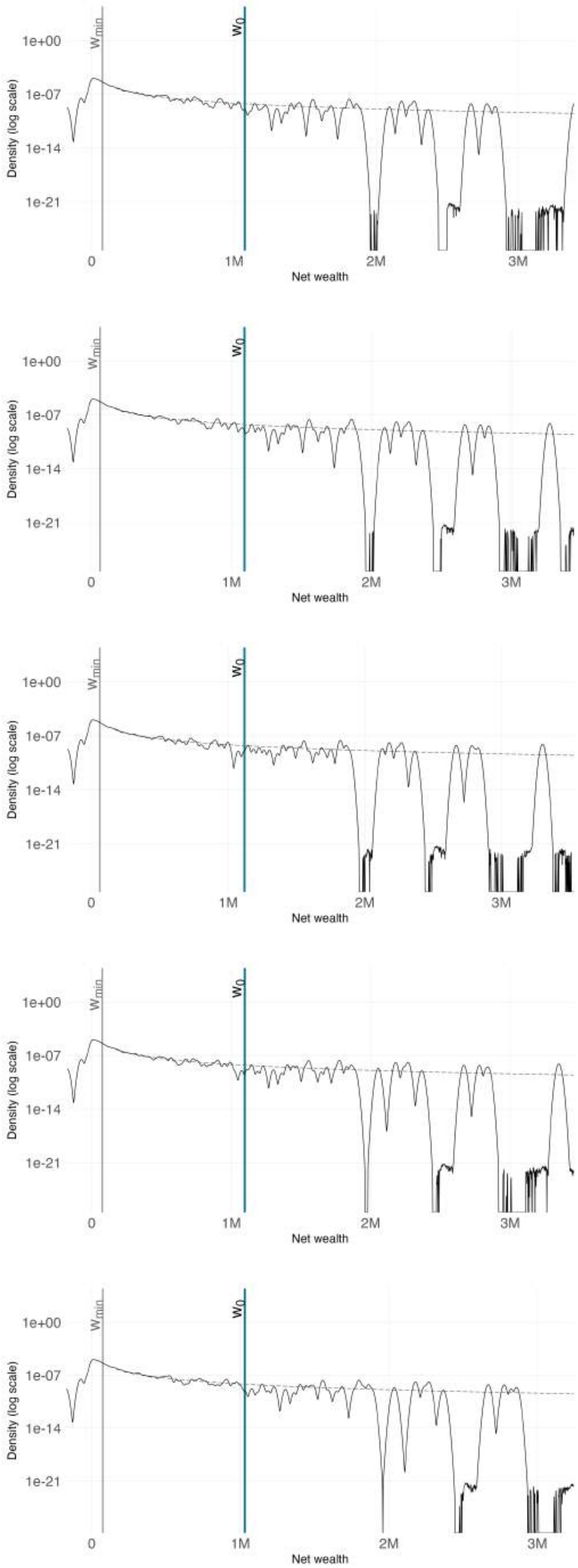


Figure D.7: Determination of Transition Threshold Parameter w_0 - IE

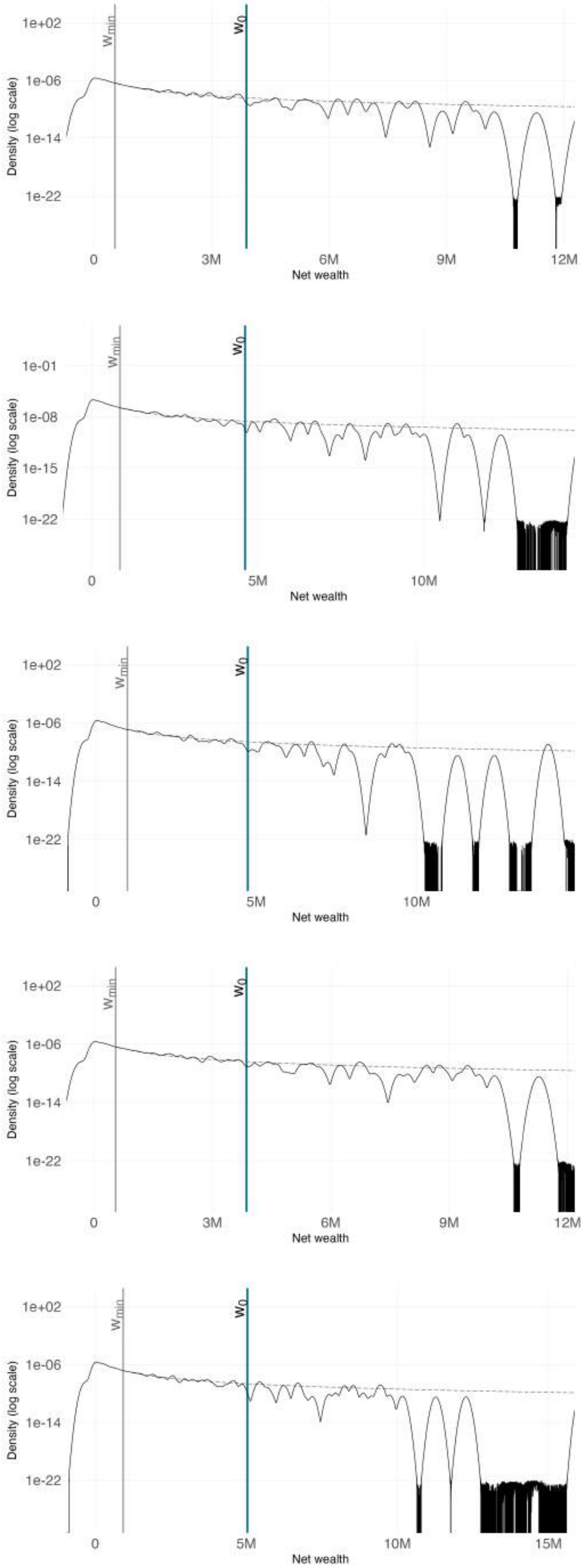


Figure D.8: Determination of Transition Threshold Parameter w_0 - IT

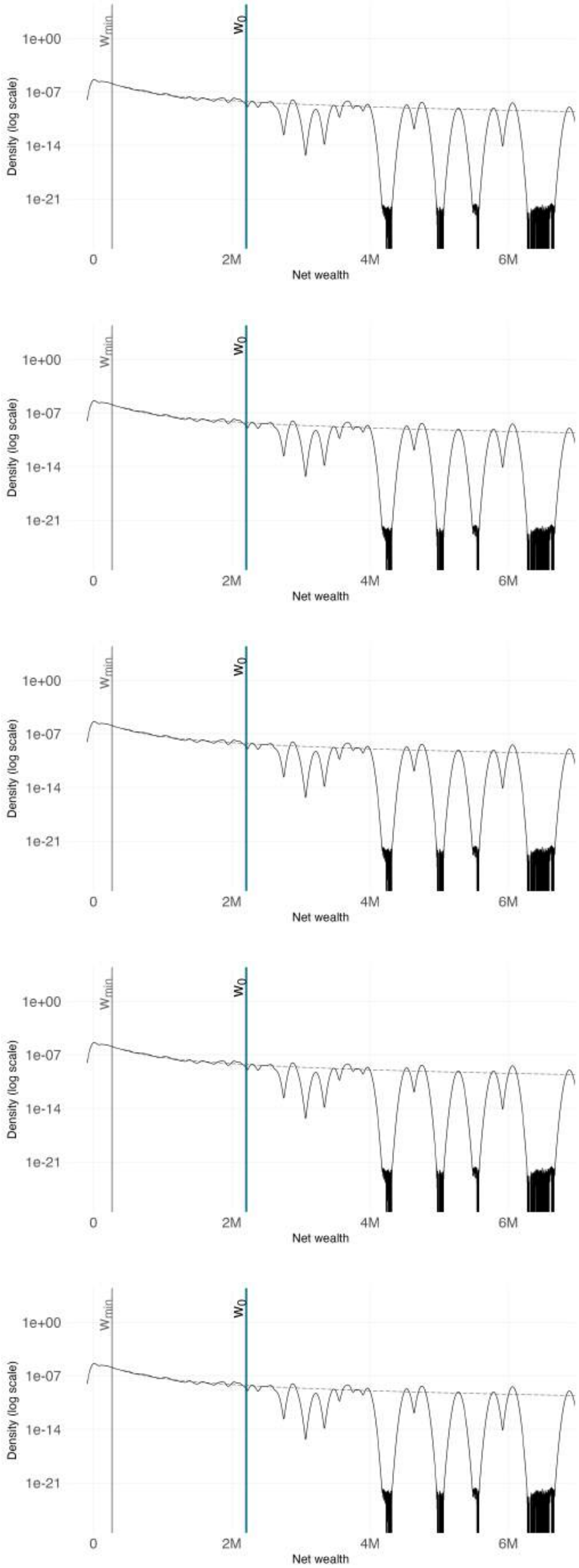


Figure D.9: Determination of Transition Threshold Parameter w_0 - LT

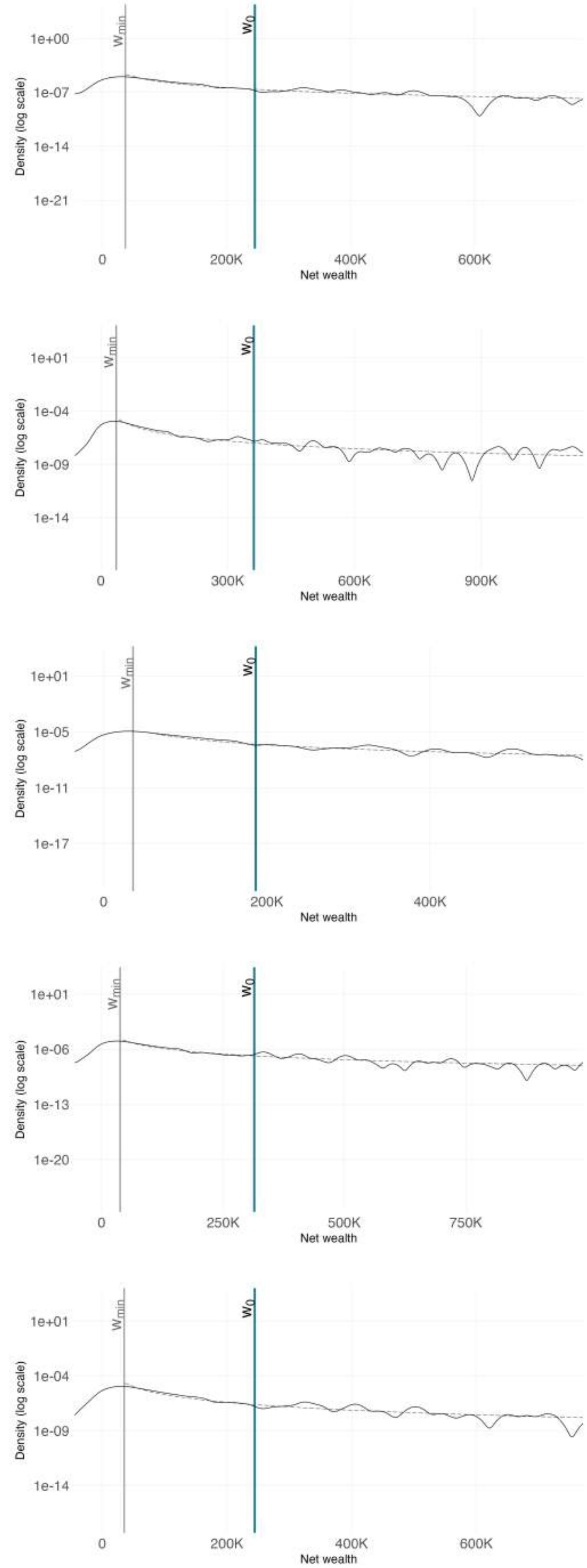


Figure D.10: Determination of Transition Threshold Parameter w_0 - LV

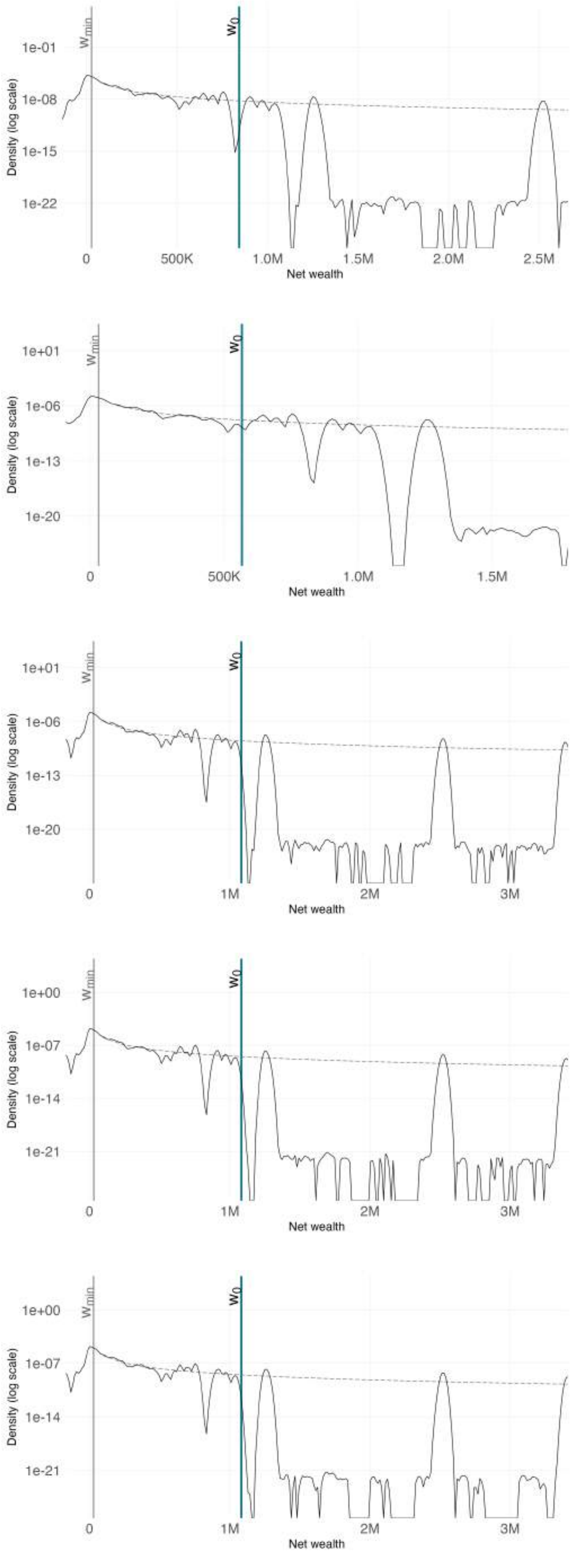


Figure D.11: Determination of Transition Threshold Parameter w_0 - NL

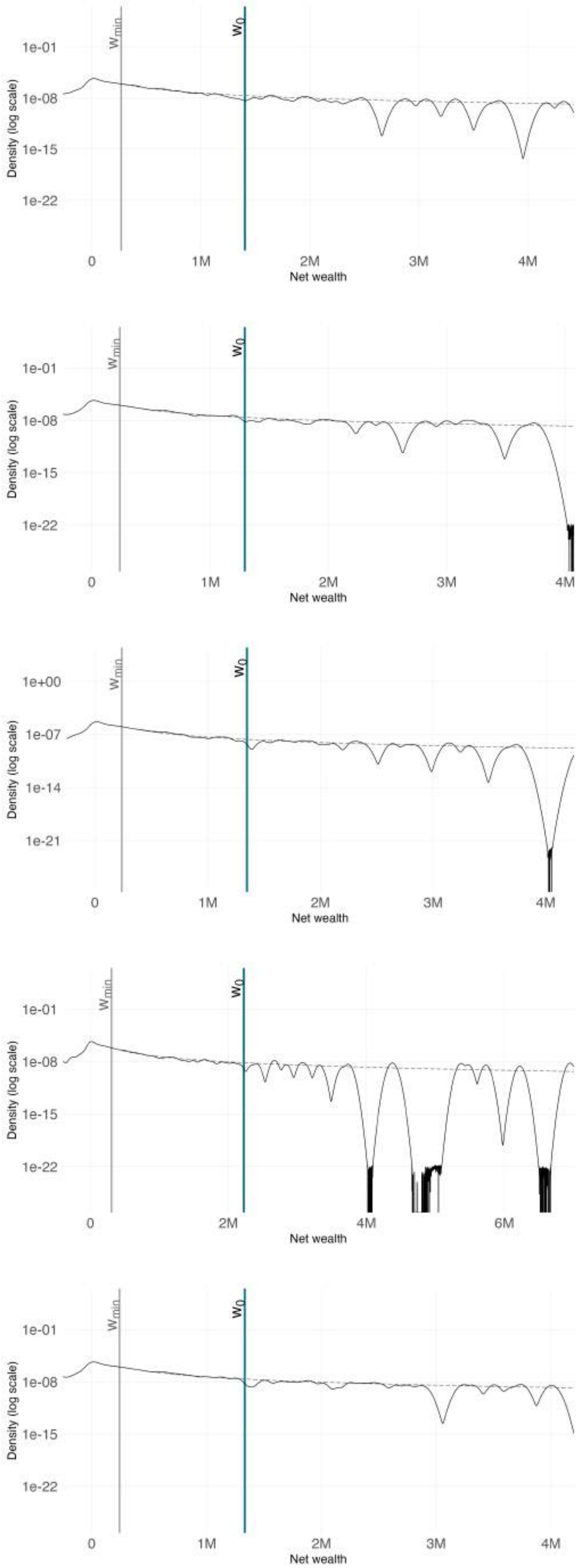


Figure D.12: Determination of Transition Threshold Parameter w_0 - PL

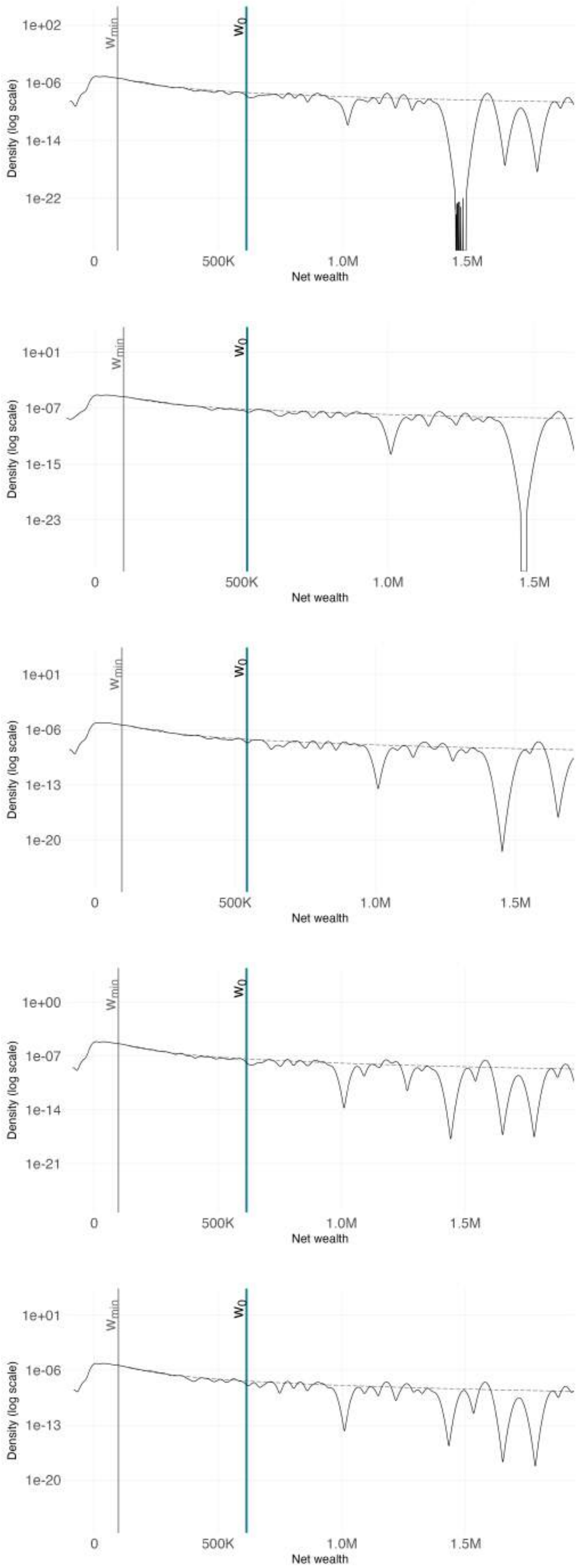


Figure D.13: Determination of Transition Threshold Parameter w_0 - PT

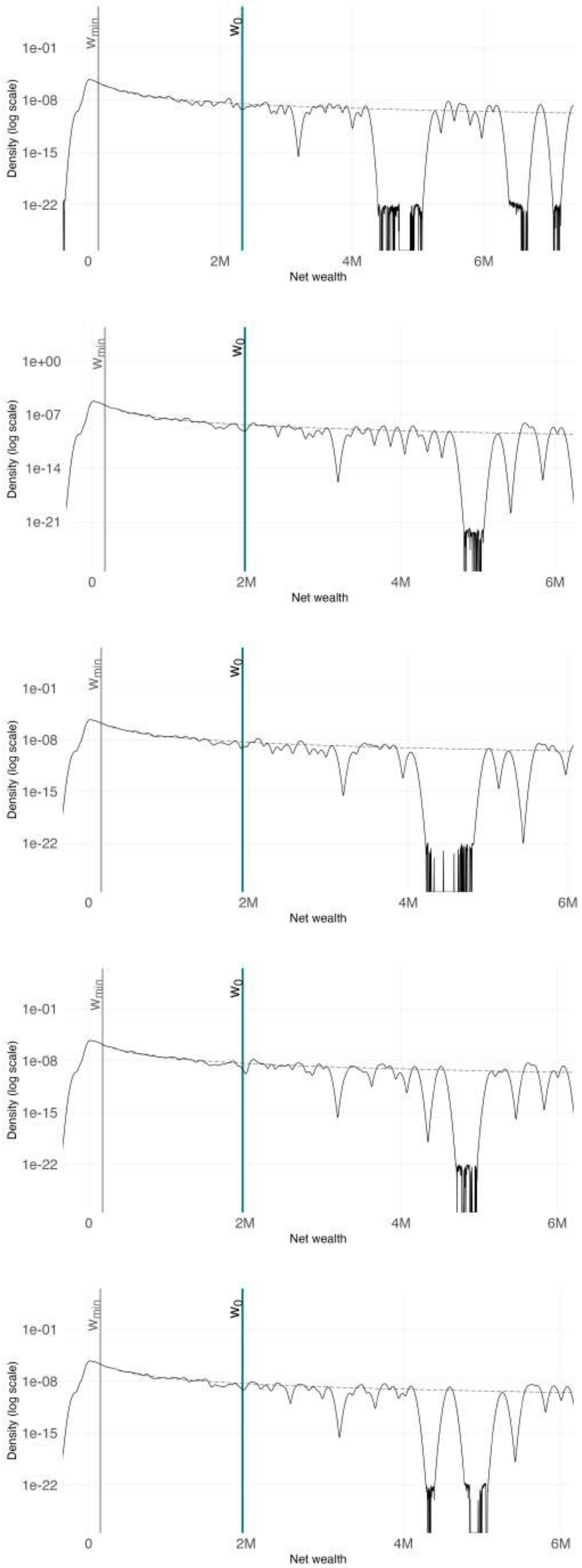
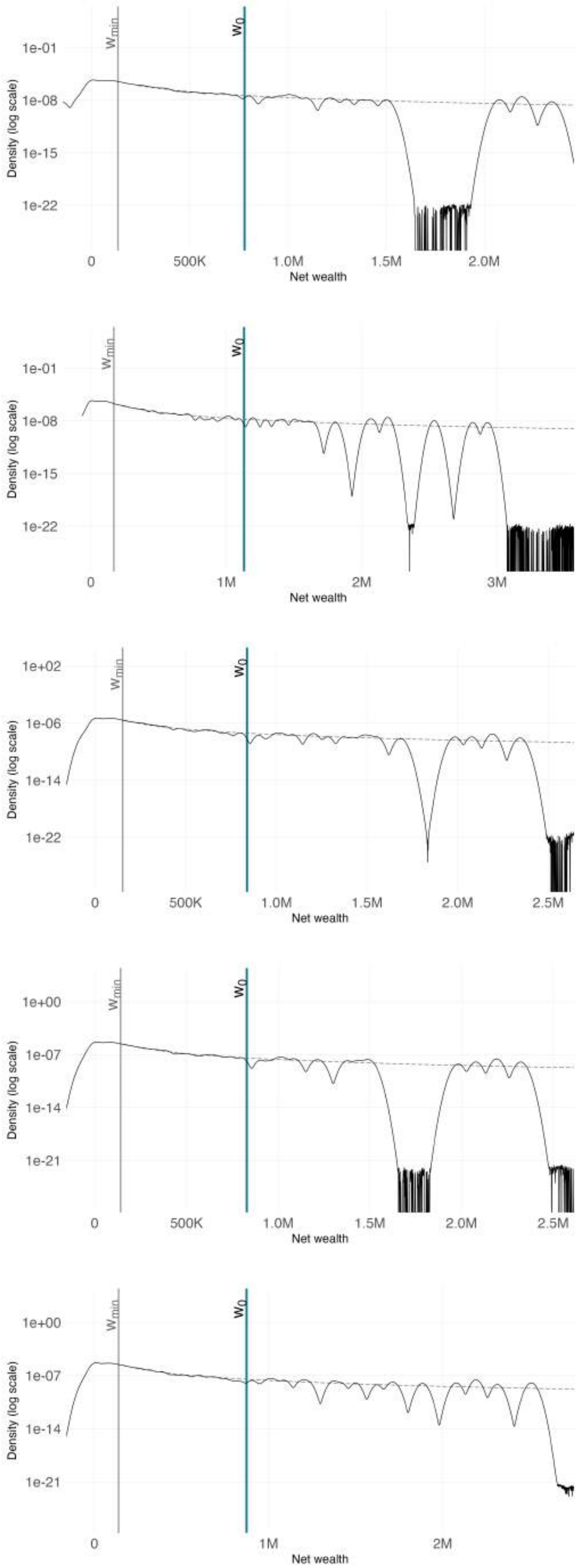


Figure D.14: Determination of Transition Threshold Parameter w_0 - SI



E Sensitivity Analysis

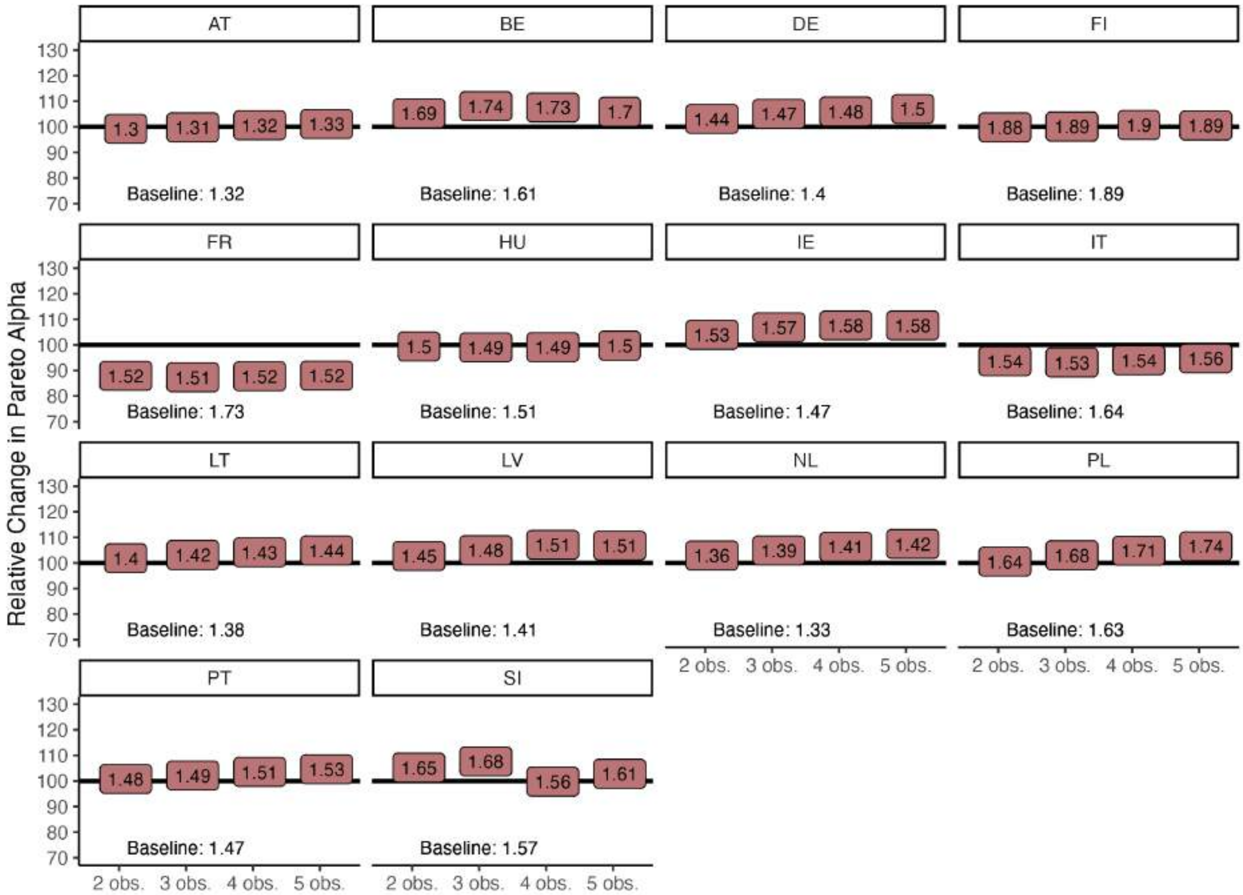
Table E.1: Overview of Sensitivity Analysis Scenarios

Overview of Sensitivity Analysis Scenarios		
A) Sensitivity towards ERLDB		
Top observations		
Drop top n	n =	1, 2, 5, 10
Drop top fraction	fraction =	0.01, 0.05, 0.1, 0.25, 0.5
Bottom observations		
Drop bottom n	n =	1, 2, 5, 10
Drop bottom fraction	fraction =	0.1, 0.25, 0.5, 0.75
Unit of observation		
Split by n	n =	2, 3, 4, 5
Reported wealth levels		
Vary wealth by constant	constant =	0.5, 0.75, 0.9, 1.1, 1.25, 1.5
B) Sensitivity towards threshold		
Arbitrary choice of w_{min}		
Fix w_{min} at percentile	percentile =	0.4, 0.5, 0.75, 0.9, 0.99
Fix w_{min} at level	level =	$2e5$, $3e5$, $5e5$, $7.5e5$, $1e6$, $1.5e6$, $2e6$
Arbitrary choice of w_0		
Fix w_0 at percentile	percentile =	0.80, 0.90, 0.95, 0.99
Fix w_0 at level	level =	$1e6$, $1.5e6$, $2e6$, $2.5e6$, $5e6$

E.1 Sensitivity Analysis: ERLDB and w_{min}

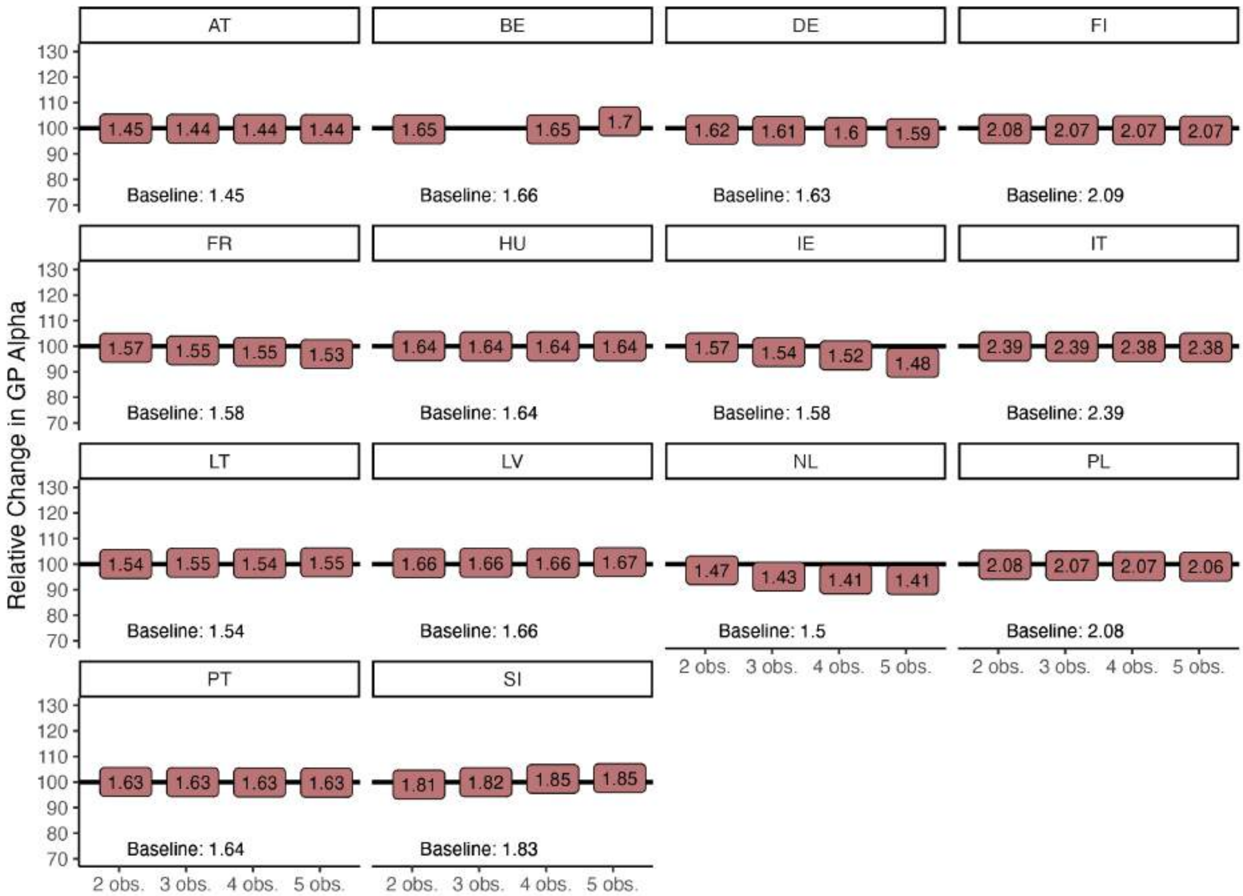
Figure E.1.	Variation in Pareto Alpha across Split by N Scenarios	99
Figure E.2.	Variation in Generalized Pareto Alpha across Split by N Scenarios	100
Figure E.3.	Variation in Pareto Alpha Across Drop Bottom N and Drop Top N Scenarios	101
Figure E.4.	Variation in Generalized Pareto Alpha Across Drop Bottom N and Drop Top N Scenarios	102
Figure E.5.	Variation in Pareto Alpha Across Drop Bottom Fraction and Drop Top Fraction Scenarios	103
Figure E.6.	Variation in Generalized Pareto Alpha Across Drop Bottom Fraction and Drop Top Fraction Scenarios	104
Figure E.7.	Variation in Pareto Alpha Across Vary Wealth by Factor Scenarios	105
Figure E.8.	Variation in Generalized Pareto Alpha Across Vary Wealth by Factor Scenarios	106
Figure E.9.	Sensitivity Analysis ERLDB and w_{min}	107
Table E.2 - Table E.15.	Sensitivity Analysis ERLDB and w_{min} by Country	108

Figure E.1: Change in Pareto α - Split by N



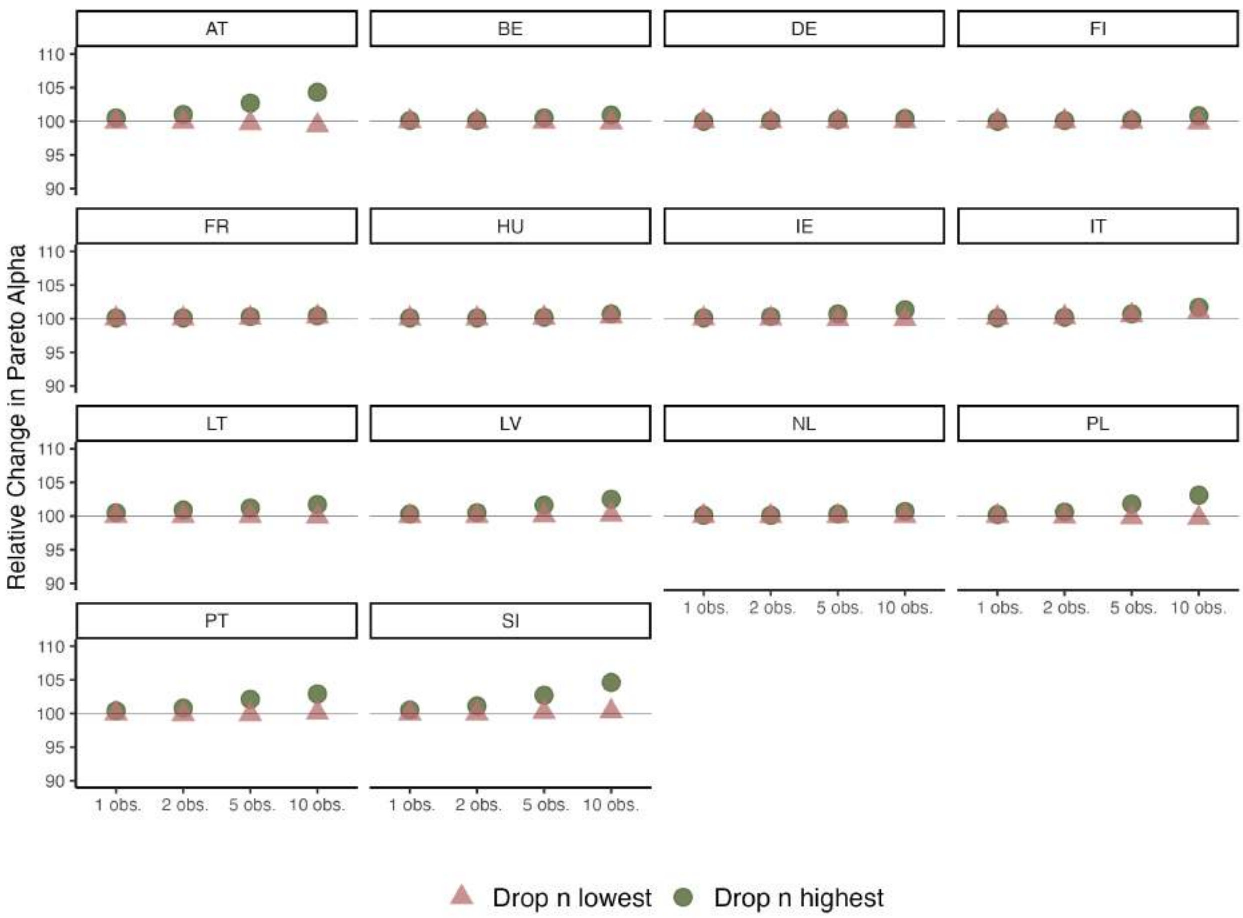
Notes: This figure shows the variation in the estimate of α, the shape parameter of the Pareto distribution, across the sensitivity scenarios *Split by n* relative to our baseline estimate. These scenarios divide the wealth of each listed observation by n to create synthetic households.

Figure E.2: Change in Generalized Pareto α_{GP} - Split by N



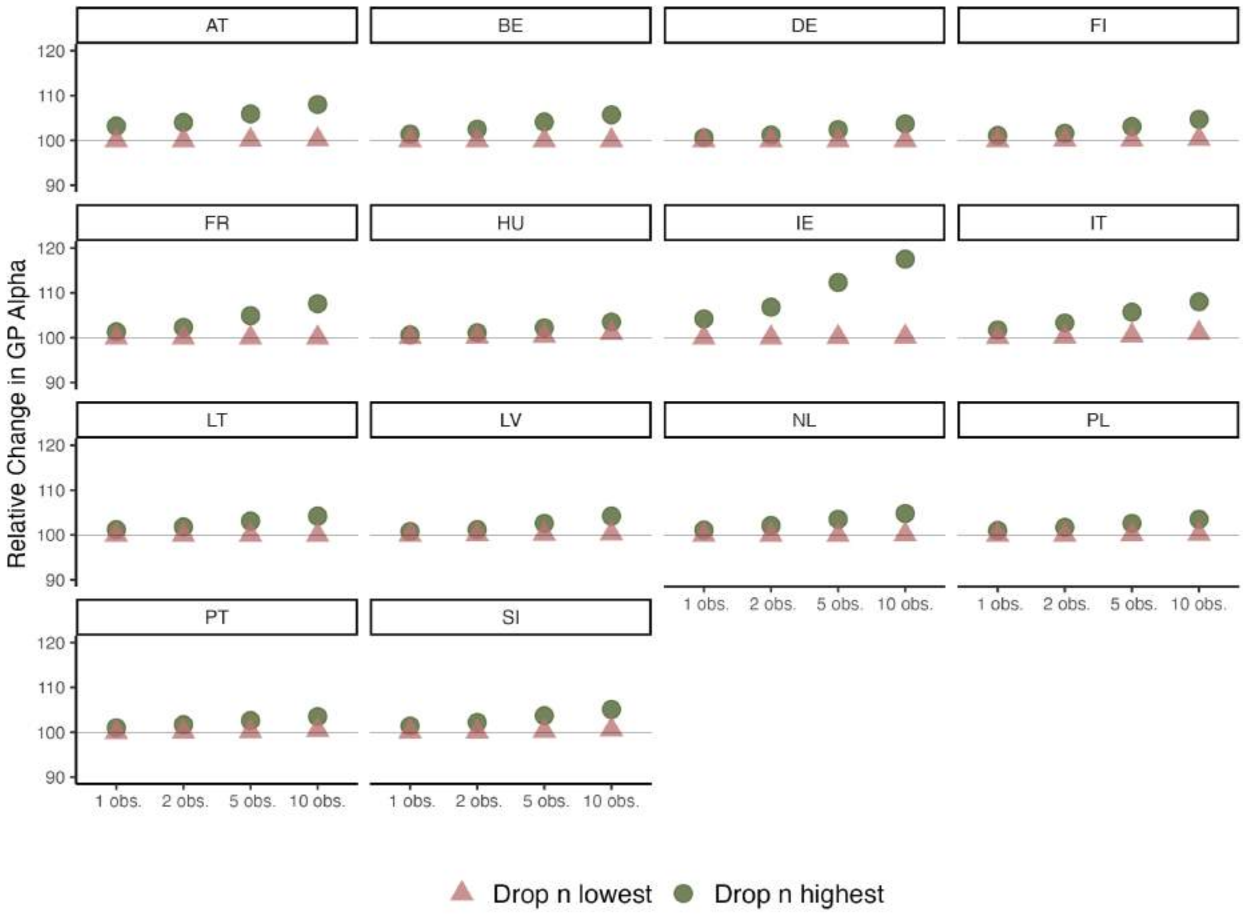
Notes: This figure shows the variation in the estimate of α_{GP}, the shape parameter of the Generalized Pareto distribution, across the sensitivity scenarios *Split by n* relative to our baseline estimate. These scenarios divide the wealth of each listed observation by n to create synthetic households.

Figure E.3: Change in Generalized Pareto α - Dropping Bottom and Top-Ranked Observations from ERLDB



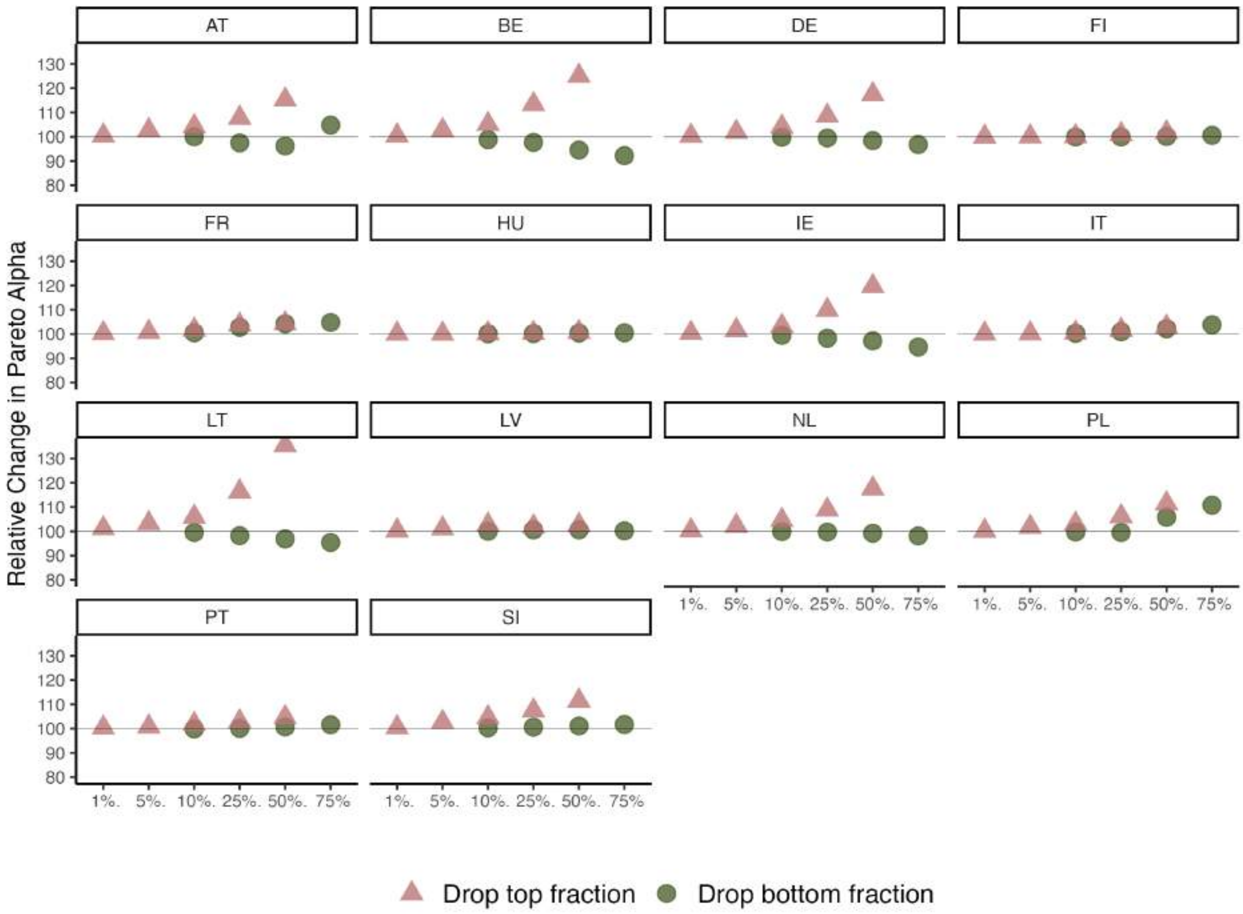
Notes: This figure shows the variation in the estimate of α, the shape parameter of the Pareto distribution, across the sensitivity scenarios *Drop bottom n* and *Drop top n* relative to our baseline estimate. These scenarios respectively omit the n bottom-ranked and top-ranked observations from each listing.

Figure E.4: Change in Generalized Pareto α - Dropping Bottom and Top-Ranked Observations from ERLDB



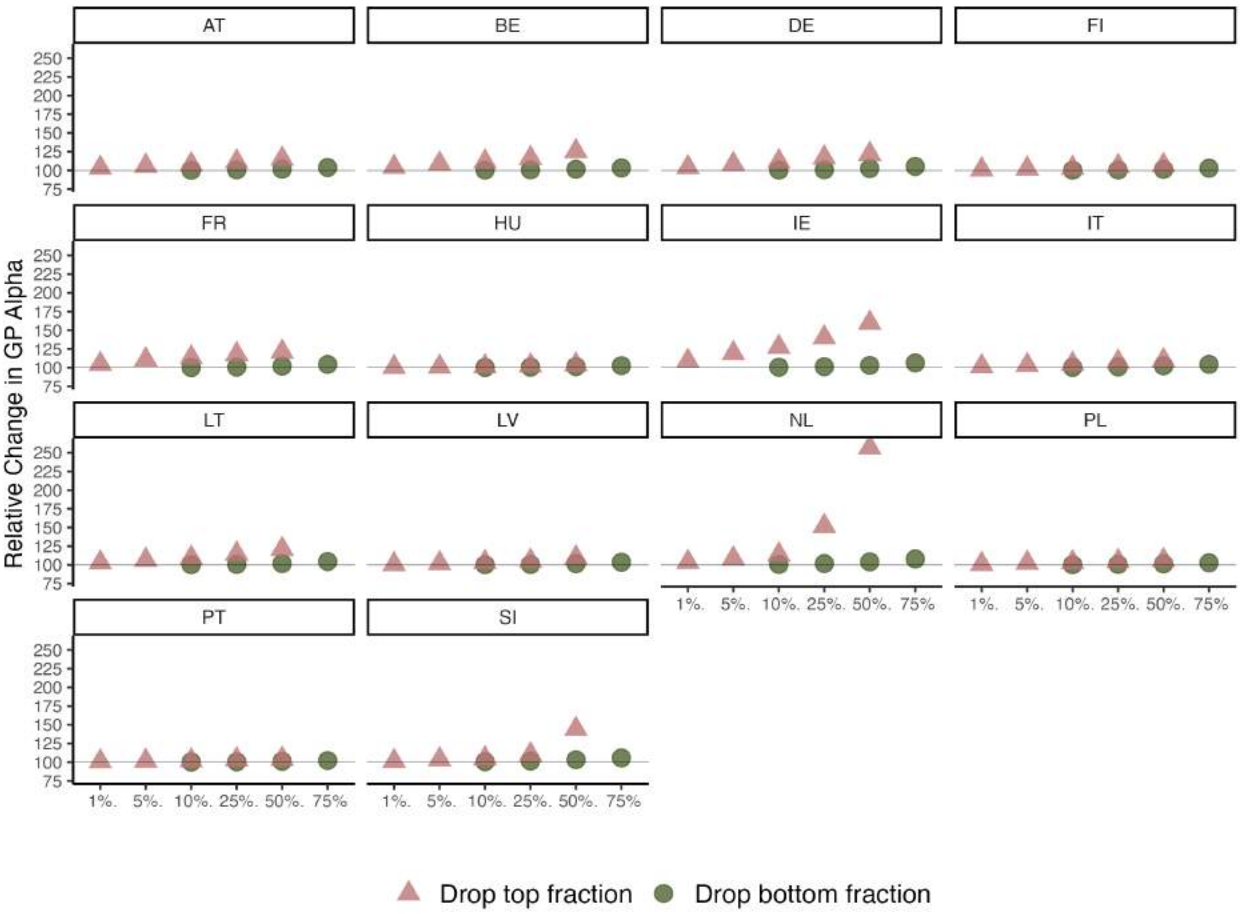
Notes: This figure shows the variation in the estimate of α_{GP}, the shape parameter of the Generalized Pareto distribution, across the sensitivity scenarios *Drop bottom n* and *Drop top n* relative to our baseline estimate. These scenarios respectively omit the n bottom-ranked and top-ranked observations from each listing.

Figure E.5: Change in Pareto α - Dropping Bottom and Top Fractions



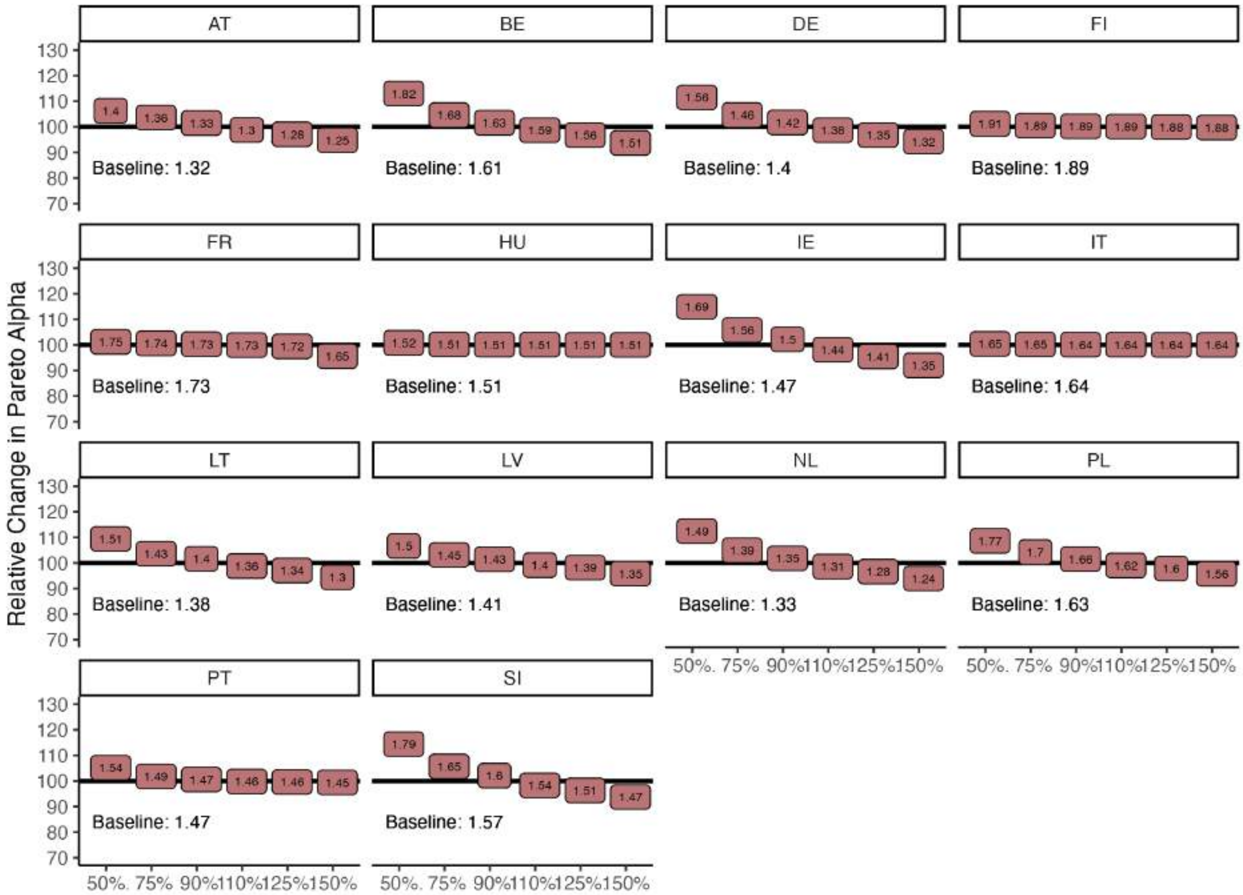
Notes: This figure shows the variation in the estimate of α, the shape parameter of the Pareto distribution, across the sensitivity scenarios *Drop bottom fraction* and *Drop top fraction* relative to our baseline estimate. These scenarios respectively omit a fraction of the bottom-ranked and top-ranked observations from each listing.

Figure E.6: Change in Pareto α - Dropping Bottom and Top Fractions



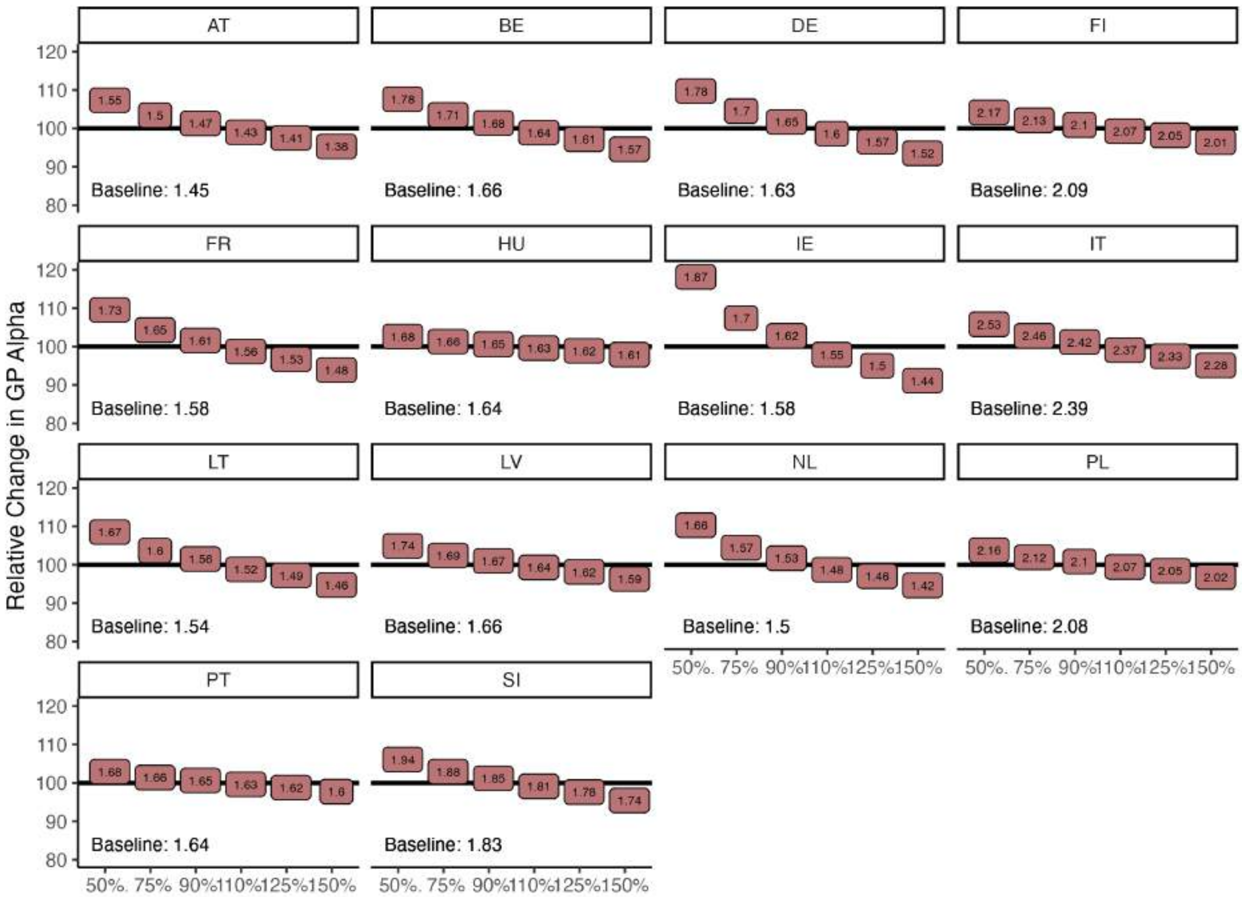
Notes: This figure shows the variation in the estimate of α_{GP}, the shape parameter of the Generalized Pareto distribution, across the sensitivity scenarios *Drop bottom fraction* and *Drop top fraction* relative to our baseline estimate. These scenarios respectively omit a fraction of the bottom-ranked and top-ranked observations from each listing.

Figure E.7: Change in Pareto α - Dropping Bottom and Top Fractions



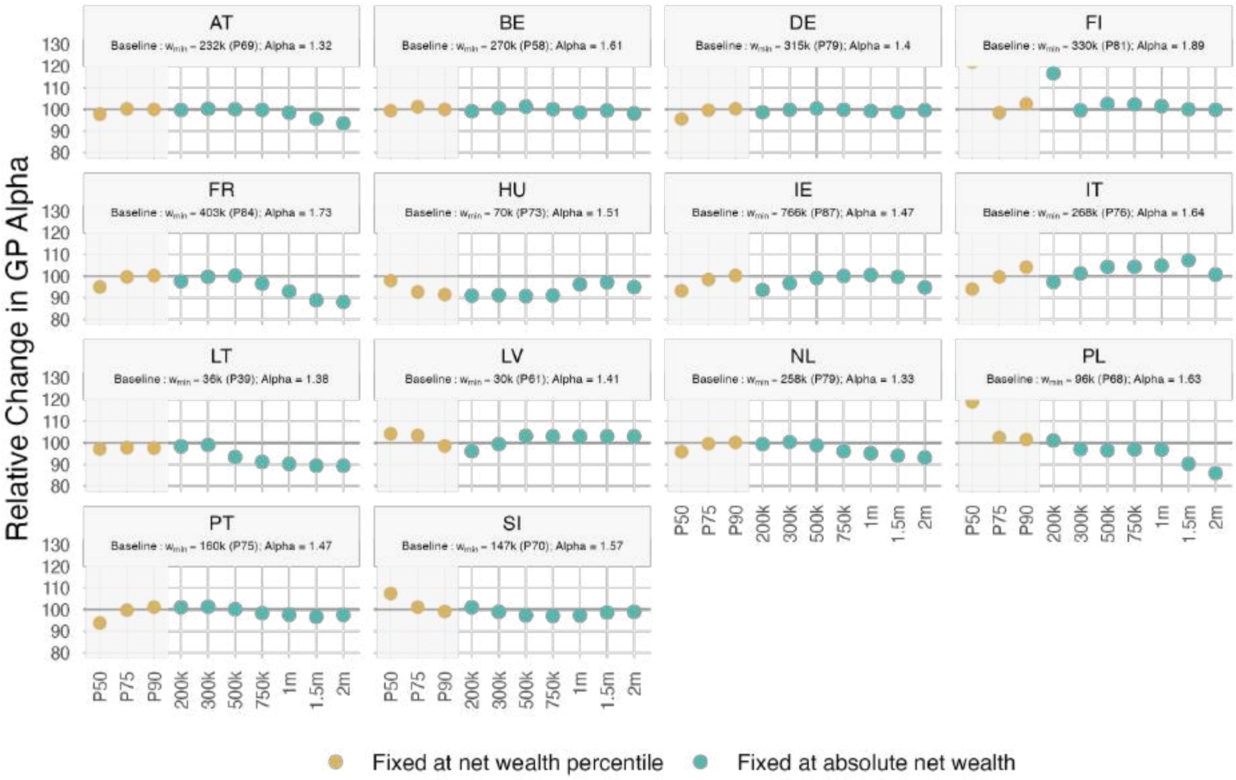
Notes: This figure shows the variation in the estimate of α, the shape parameter of the Pareto distribution, across the sensitivity scenarios *Vary wealth by constant* relative to our baseline estimate. These scenarios increase/decrease the wealth of each list observation by a constant factor.

Figure E.8: Change in Generalized Pareto α by arbitrary selection of w_{min}



Notes: This figure shows the variation in the α_{GP} parameter of the Generalized Pareto distribution across different values of w_{min}. Changes in α_{GP} are reported relative to our baseline scenario with w_{min} calculated from the RMSE minimization. The location parameters of the scenarios are set at fixed percentiles of net wealth distribution and at arbitrary absolute values of net wealth.

Figure E.9: Change in Generalized Pareto α by arbitrary selection of w_{min}



Notes: This figure shows the variation in the α_{GP} parameter of the Generalized Pareto distribution across different values of w_{min}. Changes in α_{GP} are reported relative to our baseline scenario with w_{min} calculated from the RMSE minimization. The location parameters of the scenarios are set at fixed percentiles of net wealth distribution and at arbitrary absolute values of net wealth.

Table E.2: Sensitivity Analysis: AT

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.32	39	1.45	175,183	30.7
Drop n highest						
	1	1.32	38.4	1.50	178,289	28.8
	2	1.33	37.9	1.51	178,972	28.4
	5	1.35	36.1	1.53	180,473	27.4
	10	1.37	34.5	1.56	182,099	26.5
Drop top fraction						
	0.01	1.32	38.4	1.50	178,289	28.8
	0.05	1.35	36.1	1.53	180,473	27.4
	0.10	1.37	34.5	1.56	182,099	26.5
	0.25	1.42	31.5	1.62	184,613	25.0
	0.50	1.52	26.4	1.67	186,936	23.7
Drop n lowest						
	1	1.31	39.1	1.45	175,220	30.7
	2	1.31	39.1	1.45	175,234	30.7
	5	1.31	39.4	1.45	175,297	30.6
	10	1.31	39.7	1.45	175,444	30.6
Drop bottom fraction						
	0.10	1.32	39.0	1.45	175,183	30.7
	0.25	1.28	42.0	1.46	175,915	30.4
	0.50	1.27	43.8	1.47	177,243	29.8
	0.75	1.38	34.1	1.50	179,308	28.5
Split by n						
	2	1.30	40.1	1.45	174,045	30.8
	3	1.31	39.3	1.44	173,637	30.8
	4	1.32	38.4	1.44	172,750	30.8
	5	1.33	37.9	1.44	172,776	30.8
Vary wealth by factor						
	0.50	1.40	32.8	1.55	181,304	26.8
	0.75	1.36	35.3	1.50	178,261	28.8
	0.90	1.33	37.4	1.47	176,293	29.9
	1.00	1.32	39.0	1.45	175,183	30.7
	1.10	1.30	40.5	1.43	174,189	31.4
	1.25	1.28	42.7	1.41	172,863	32.5
	1.50	1.25	45.9	1.38	170,833	34.2
Fix wmin at level						
	200,000	1.31	39.4	1.47	161,708	30.1
	300,000	1.30	40.1	1.42	206,332	31.5
	500,000	1.25	45.3	1.41	341,941	31.8
	750,000	1.23	48.7	1.31	401,278	32.3
	1,000,000	1.21	52.2	1.26	487,958	32.6
	1,500,000	1.16	59.9	1.19	648,239	33.5
	2,000,000	1.14	64.9	1.22	1,347,733	38.2
Fix wmin at percentile						
	0.40	1.22	49.4	2.42	257,510	18.0
	0.50	1.25	45.6	2.04	233,923	21.3
	0.75	1.30	39.6	1.43	193,954	31.2
	0.90	1.25	45.6	1.40	355,027	31.7
	0.99	1.13	66.4	1.23	1,547,598	40.5

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.3: Sensitivity Analysis: BE

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.61	21.3	1.66	207,552	21.2
Drop n highest						
	1	1.61	21.2	1.68	209,061	20.7
	2	1.61	21.2	1.70	209,982	20.3
	5	1.62	21.0	1.73	211,445	19.8
	10	1.62	20.8	1.75	212,905	19.3
Drop top fraction						
	0.01	1.62	21.0	1.73	211,764	19.7
	0.05	1.65	19.9	1.81	215,946	18.4
	0.10	1.69	18.8	1.85	219,045	17.8
	0.25	1.82	15.9	1.93	227,207	16.9
	0.50	2.01	12.9	2.08	248,177	15.6
Drop n lowest						
	1	1.61	21.3	1.66	207,553	21.2
	2	1.61	21.3	1.66	207,566	21.2
	5	1.61	21.3	1.66	207,586	21.2
	10	1.61	21.4	1.66	207,603	21.2
Drop bottom fraction						
	0.10	1.59	21.9	1.66	208,010	21.1
	0.25	1.57	22.5	1.67	208,871	21.0
	0.50	1.52	24.6	1.68	210,716	20.7
	0.75	1.48	26.2	1.71	213,877	20.2
Split by n						
	2	1.69	18.8	1.65	205,202	21.3
	3	1.74	17.7	Inf	0.0	2.95
	4	1.73	17.9	1.65	204,703	21.3
	5	1.70	18.6	1.70	217,728	20.6
Vary wealth by factor						
	0.50	1.82	16.0	1.78	215,130	18.8
	0.75	1.68	19.1	1.71	210,913	20.0
	0.90	1.63	20.5	1.68	208,957	20.7
	1.00	1.61	21.3	1.66	207,552	21.2
	1.10	1.59	22.0	1.64	206,166	21.7
	1.25	1.56	23.2	1.61	204,279	22.4
	1.50	1.51	25.2	1.57	201,197	23.5
Fix wmin at level						
	200,000	1.59	22.4	1.78	201,817	19.8
	300,000	1.61	21.1	1.65	220,740	21.4
	500,000	1.63	20.5	1.53	286,258	22.7
	750,000	1.62	21.6	1.47	400,969	23.2
	1,000,000	1.59	23.2	1.58	719,650	23.2
	1,500,000	0.94	129.4	1.20	303,378	15.4
	2,000,000	0.89	159.8	1.15	237,843	11.3
Fix wmin at percentile						
	0.40	1.57	23.4	1.94	212,807	18.6
	0.50	1.59	22.1	1.74	198,896	20.2
	0.75	1.63	20.4	1.56	255,923	22.3
	0.90	1.62	21.7	1.47	409,512	23.2
	0.99	0.86	178.9	1.13	36,581	NaN

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.4: Sensitivity Analysis: DE

	Scenario	Pareto		GPareto		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.4	32.9	1.63	218,406	24.6
Drop n highest						
	1	1.4	32.9	1.64	219,360	24.4
	2	1.4	32.9	1.65	220,167	24.2
	5	1.4	32.7	1.67	222,083	23.8
	10	1.4	32.6	1.69	223,803	23.4
Drop top fraction						
	0.01	1.41	32.5	1.69	224,128	23.3
	0.05	1.43	31.2	1.76	231,214	22.1
	0.10	1.46	29.5	1.82	238,001	21.4
	0.25	1.52	26.5	1.91	247,379	20.5
	0.50	1.65	21.8	1.98	251,936	19.6
Drop n lowest						
	1	1.4	32.9	1.63	218,408	24.6
	2	1.4	32.9	1.63	218,410	24.6
	5	1.4	32.9	1.63	218,434	24.6
	10	1.4	32.9	1.63	218,449	24.6
Drop bottom fraction						
	0.10	1.40	33.1	1.63	218,983	24.5
	0.25	1.39	33.4	1.64	220,524	24.4
	0.50	1.38	34.4	1.67	224,200	23.9
	0.75	1.35	36.1	1.71	229,689	23.1
Split by n						
	2	1.44	30.4	1.62	214,178	24.7
	3	1.47	28.9	1.61	211,437	24.7
	4	1.48	28.2	1.60	208,702	24.8
	5	1.50	27.5	1.59	205,356	24.8
Vary wealth by factor						
	0.50	1.56	24.8	1.78	233,170	21.8
	0.75	1.46	29.2	1.70	225,483	23.2
	0.90	1.42	31.5	1.65	221,113	24.1
	1.00	1.40	32.9	1.63	218,406	24.6
	1.10	1.38	34.2	1.60	215,620	25.1
	1.25	1.35	36.1	1.57	211,463	25.9
	1.50	1.32	39.1	1.52	206,116	27.2
Fix wmin at level						
	200,000	1.39	33.7	1.84	221,909	22.8
	300,000	1.40	33.0	1.64	213,898	24.5
	500,000	1.39	34.1	1.55	292,575	25.4
	750,000	1.36	36.4	1.52	434,115	25.6
	1,000,000	1.35	38.0	1.41	475,260	26.0
	1,500,000	1.34	39.7	1.28	403,626	20.3
	2,000,000	1.19	56.0	1.25	433,656	20.1
Fix wmin at percentile						
	0.40	1.36	38.3	2.27	219,766	19.4
	0.50	1.37	35.8	2.18	222,769	20.1
	0.75	1.40	33.1	1.69	217,957	24.3
	0.90	1.38	34.3	1.56	333,193	25.4
	0.99	1.10	73.2	1.24	580,310	22.6

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.5: Sensitivity Analysis: FI

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.89	15.5	2.09	201,258	15.2
Drop n highest						
	1	1.89	15.5	2.11	201,649	15.1
	2	1.89	15.4	2.12	201,964	15.0
	5	1.89	15.4	2.15	202,376	14.8
	10	1.90	15.2	2.19	202,455	14.5
Drop top fraction						
	0.01	1.89	15.5	2.11	201,649	15.1
	0.05	1.89	15.4	2.13	202,234	14.9
	0.10	1.89	15.4	2.15	202,376	14.8
	0.25	1.90	15.2	2.20	202,714	14.4
	0.50	1.92	14.9	2.23	202,818	14.2
Drop n lowest						
	1	1.89	15.5	2.09	201,270	15.2
	2	1.89	15.5	2.09	201,307	15.2
	5	1.88	15.5	2.09	201,339	15.2
	10	1.88	15.5	2.09	201,496	15.2
Drop bottom fraction						
	0.10	1.88	15.5	2.09	201,339	15.2
	0.25	1.88	15.5	2.10	201,567	15.2
	0.50	1.89	15.4	2.11	202,230	15.1
	0.75	1.90	15.3	2.15	203,072	14.8
Split by n						
	2	1.88	15.5	2.08	199,898	15.3
	3	1.89	15.5	2.07	199,042	15.3
	4	1.90	15.3	2.07	198,550	15.3
	5	1.89	15.3	2.07	198,136	15.3
Vary wealth by factor						
	0.50	1.91	15.1	2.17	202,785	14.6
	0.75	1.89	15.3	2.13	202,179	14.9
	0.90	1.89	15.4	2.10	201,683	15.1
	1.00	1.89	15.5	2.09	201,258	15.2
	1.10	1.89	15.5	2.07	200,815	15.4
	1.25	1.88	15.5	2.05	200,216	15.6
	1.50	1.88	15.6	2.01	199,324	15.9
Fix wmin at level						
	200,000	1.72	19.0	2.57	193,235	13.7
	300,000	1.87	15.7	2.17	197,645	14.9
	500,000	1.95	14.9	1.91	252,600	15.8
	750,000	1.91	16.2	1.86	370,494	15.8
	1,000,000	1.77	19.6	1.77	471,300	15.9
	1,500,000	1.61	25.1	1.76	732,741	16.2
	2,000,000	1.58	26.5	1.92	1,143,542	16.1
Fix wmin at percentile						
	0.40	1.35	37.9	3.43	191,076	12.1
	0.50	1.52	26.5	2.94	182,411	13.0
	0.75	1.82	16.5	2.31	195,640	14.5
	0.90	1.95	14.9	1.91	246,742	15.7
	0.99	1.60	25.5	1.88	895,580	16.9

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.6: Sensitivity Analysis: FR

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.73	18.7	1.58	224,855	22
Drop n highest						
	1	1.73	18.7	1.60	226,376	21.6
	2	1.73	18.7	1.62	227,564	21.3
	5	1.73	18.6	1.66	230,461	20.5
	10	1.74	18.5	1.70	233,222	19.8
Drop top fraction						
	0.01	1.73	18.6	1.66	230,461	20.5
	0.05	1.74	18.4	1.75	236,365	19.0
	0.10	1.76	18.0	1.80	239,064	18.4
	0.25	1.80	17.2	1.86	242,822	17.7
	0.50	1.81	17.0	1.91	245,252	17.1
Drop n lowest						
	1	1.73	18.7	1.58	224,860	22
	2	1.73	18.7	1.58	224,881	22
	5	1.73	18.7	1.58	224,892	22
	10	1.73	18.6	1.58	224,927	22
Drop bottom fraction						
	0.10	1.74	18.5	1.58	225,231	22.0
	0.25	1.78	17.6	1.59	226,235	21.8
	0.50	1.80	17.1	1.61	229,139	21.5
	0.75	1.81	16.9	1.65	233,285	20.8
Split by n						
	2	1.52	25.6	1.57	221,504	22.1
	3	1.51	26.1	1.55	215,998	22.3
	4	1.52	25.8	1.55	213,213	22.4
	5	1.52	25.6	1.53	208,386	22.5
Vary wealth by factor						
	0.50	1.75	18.3	1.73	235,111	19.4
	0.75	1.74	18.5	1.65	229,689	20.7
	0.90	1.73	18.6	1.61	226,667	21.5
	1.00	1.73	18.7	1.58	224,855	22.0
	1.10	1.73	18.8	1.56	223,089	22.6
	1.25	1.72	19.0	1.53	220,605	23.3
	1.50	1.65	20.8	1.48	217,038	24.5
Fix wmin at level						
	200,000	1.56	23.8	1.84	182,289	19.8
	300,000	1.67	20.2	1.71	202,800	20.9
	500,000	1.51	26.5	1.48	231,210	22.3
	750,000	1.40	33.2	1.33	272,177	22.8
	1,000,000	1.36	36.9	1.31	392,021	22.9
	1,500,000	1.31	41.6	1.29	742,516	23.5
	2,000,000	1.27	46.0	1.25	914,993	22.3
Fix wmin at percentile						
	0.40	1.44	30.4	2.45	199,132	16.2
	0.50	1.50	27.2	2.11	184,984	18.0
	0.75	1.66	20.3	1.71	200,839	20.9
	0.90	1.48	27.9	1.43	232,597	22.5
	0.99	1.30	43.2	1.27	856,427	22.5

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.7: Sensitivity Analysis: HU

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.51	24.6	1.64	54,013	22.3
Drop n highest						
	1	1.51	24.6	1.65	54,102	22.1
	2	1.51	24.6	1.66	54,196	21.9
	5	1.51	24.5	1.68	54,357	21.5
	10	1.52	24.2	1.70	54,544	21.1
Drop top fraction						
	0.01	1.51	24.6	1.65	54,102	22.1
	0.05	1.51	24.6	1.66	54,196	21.9
	0.10	1.51	24.5	1.67	54,264	21.7
	0.25	1.51	24.3	1.69	54,447	21.3
	0.50	1.52	24.1	1.71	54,639	20.9
Drop n lowest						
	1	1.51	24.6	1.64	54,027	22.3
	2	1.51	24.6	1.64	54,039	22.2
	5	1.51	24.6	1.65	54,092	22.1
	10	1.51	24.4	1.66	54,185	21.9
Drop bottom fraction						
	0.10	1.51	24.6	1.65	54,055	22.2
	0.25	1.51	24.5	1.65	54,128	22.1
	0.50	1.51	24.4	1.66	54,258	21.8
	0.75	1.52	24.3	1.69	54,458	21.3
Split by n						
	2	1.50	25.1	1.64	53,909	22.3
	3	1.49	25.4	1.64	53,837	22.3
	4	1.49	25.3	1.64	53,782	22.3
	5	1.50	24.9	1.64	53,739	22.3
Vary wealth by factor						
	0.50	1.52	24.2	1.68	54,392	21.4
	0.75	1.51	24.6	1.66	54,212	21.8
	0.90	1.51	24.6	1.65	54,083	22.1
	1.00	1.51	24.6	1.64	54,013	22.3
	1.10	1.51	24.6	1.63	53,937	22.5
	1.25	1.51	24.7	1.62	53,831	22.7
	1.50	1.51	24.7	1.61	53,655	23.2
Fix wmin at level						
	200,000	1.47	28.9	1.59	138,066	23.2
	300,000	1.38	35.1	1.56	195,625	23.4
	500,000	1.34	39.2	1.63	363,992	24.4
	750,000	1.33	41.0	1.89	688,156	24.8
	1,000,000	1.33	41.2	1.77	706,641	16.8
	1,500,000	1.32	43.2	1.63	786,579	NaN
	2,000,000	1.24	51.4	1.67	1,382,581	NaN
Fix wmin at percentile						
	0.40	1.35	33.3	1.80	38,342	20.5
	0.50	1.40	29.8	1.72	39,724	21.4
	0.75	1.51	24.4	1.64	55,103	22.3
	0.90	1.53	25.3	1.52	86,316	23.5
	0.99	1.34	40.0	1.65	439,514	23.4

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.8: Sensitivity Analysis: IE

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.47	28.2	1.58	499,495	24.3
Drop n highest						
	1	1.48	28.1	1.65	507,192	22.9
	2	1.48	28.0	1.69	511,731	22.1
	5	1.48	27.7	1.78	520,130	20.8
	10	1.49	27.2	1.86	527,133	19.7
Drop top fraction						
	0.01	1.48	27.9	1.72	515,117	21.6
	0.05	1.50	27.0	1.88	528,919	19.4
	0.10	1.52	25.9	2.02	543,215	18.2
	0.25	1.62	22.2	2.22	561,448	16.8
	0.50	1.76	18.3	2.53	596,482	15.6
Drop n lowest						
	1	1.47	28.2	1.58	499,531	24.3
	2	1.47	28.2	1.58	499,720	24.3
	5	1.47	28.2	1.58	499,774	24.3
	10	1.47	28.3	1.59	500,750	24.2
Drop bottom fraction						
	0.10	1.47	28.5	1.59	501,808	24.1
	0.25	1.45	29.5	1.60	504,765	23.9
	0.50	1.43	30.4	1.63	510,416	23.4
	0.75	1.40	32.8	1.68	517,800	22.4
Split by n						
	2	1.53	25.6	1.57	490,408	24.3
	3	1.57	23.8	1.54	462,041	24.3
	4	1.58	23.5	1.52	442,034	24.3
	5	1.58	23.5	1.48	395,968	24.3
Vary wealth by factor						
	0.50	1.69	20.1	1.87	531,077	19.7
	0.75	1.56	24.4	1.70	513,666	22.0
	0.90	1.50	26.7	1.62	505,568	23.4
	1.00	1.47	28.2	1.58	499,495	24.3
	1.10	1.44	29.7	1.55	495,430	25.1
	1.25	1.41	32.0	1.50	489,142	26.4
	1.50	1.35	35.6	1.44	480,617	28.4
Fix wmin at level						
	200,000	1.37	33.9	1.75	262,995	22.7
	300,000	1.43	29.2	1.71	308,838	23.2
	500,000	1.47	27.7	1.65	389,088	23.9
	750,000	1.48	28.0	1.61	518,328	24.2
	1,000,000	1.47	29.0	1.53	593,922	24.8
	1,500,000	1.40	34.2	1.27	470,544	23.8
	2,000,000	1.31	41.5	1.20	590,678	24.1
Fix wmin at percentile						
	0.40	1.32	38.6	1.81	238,528	22.1
	0.50	1.36	34.4	1.76	257,260	22.6
	0.75	1.46	28.1	1.66	344,937	23.7
	0.90	1.48	28.2	1.58	554,762	24.4
	0.99	1.19	56.5	1.19	1,272,336	29.7

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.9: Sensitivity Analysis: IT

	Scenario	Pareto		GPareto		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.64	19.9	2.39	201,717	13.3
Drop n highest						
	1	1.65	19.9	2.44	202,236	13.1
	2	1.65	19.8	2.47	202,741	12.9
	5	1.66	19.6	2.53	203,426	12.6
	10	1.67	19.1	2.59	204,282	12.4
Drop top fraction						
	0.01	1.65	19.9	2.44	202,236	13.1
	0.05	1.65	19.8	2.47	202,741	12.9
	0.10	1.65	19.7	2.52	203,254	12.7
	0.25	1.67	19.1	2.58	204,165	12.4
	0.50	1.69	18.6	2.63	204,879	12.2
Drop n lowest						
	1	1.65	19.9	2.40	201,769	13.3
	2	1.65	19.9	2.40	201,790	13.3
	5	1.65	19.7	2.41	201,906	13.2
	10	1.66	19.5	2.42	202,070	13.2
Drop bottom fraction						
	0.10	1.65	19.8	2.40	201,809	13.2
	0.25	1.66	19.5	2.42	202,028	13.2
	0.50	1.68	18.9	2.45	202,519	13.0
	0.75	1.71	18.2	2.50	203,152	12.8
Split by n						
	2	1.54	23.6	2.39	201,390	13.3
	3	1.53	24.0	2.39	201,219	13.3
	4	1.54	23.4	2.38	200,922	13.3
	5	1.56	22.9	2.38	200,647	13.3
Vary wealth by factor						
	0.50	1.65	19.8	2.53	203,440	12.6
	0.75	1.65	19.9	2.46	202,541	13.0
	0.90	1.64	19.9	2.42	202,070	13.1
	1.00	1.64	19.9	2.39	201,717	13.3
	1.10	1.64	20.0	2.37	201,339	13.4
	1.25	1.64	20.0	2.33	200,776	13.6
	1.50	1.64	20.0	2.28	199,737	13.9
Fix wmin at level						
	200,000	1.53	24.1	2.34	168,374	13.4
	300,000	1.67	19.4	2.34	210,427	13.5
	500,000	1.44	29.7	2.10	271,101	14.1
	750,000	1.39	33.4	2.06	395,185	14.2
	1,000,000	1.37	36.0	2.17	548,149	14.6
	1,500,000	1.33	40.0	1.68	529,631	13.8
	2,000,000	1.28	45.5	1.46	626,755	15.9
Fix wmin at percentile						
	0.40	1.38	33.1	2.95	161,316	11.8
	0.50	1.44	28.7	2.54	153,957	12.8
	0.75	1.64	20.1	2.41	199,988	13.2
	0.90	1.44	29.3	2.10	261,789	14.0
	0.99	1.31	42.0	1.52	437,652	13.1

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.10: Sensitivity Analysis: LT

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.38	29.9	1.54	37,053	24.8
Drop n highest						
	1	1.38	29.4	1.56	37,284	24.2
	2	1.39	29.1	1.57	37,428	23.9
	5	1.39	28.8	1.59	37,747	23.3
	10	1.40	28.4	1.61	38,071	22.8
Drop top fraction						
	0.01	1.39	28.8	1.59	37,747	23.3
	0.05	1.42	27.1	1.64	38,584	21.9
	0.10	1.46	25.1	1.69	39,305	20.8
	0.25	1.60	19.4	1.77	40,439	19.2
	0.50	1.87	13.4	1.87	42,143	17.6
Drop n lowest						
	1	1.38	29.9	1.54	37,053	24.8
	2	1.38	29.9	1.54	37,054	24.8
	5	1.38	29.9	1.54	37,054	24.8
	10	1.38	30.0	1.54	37,068	24.8
Drop bottom fraction						
	0.10	1.37	30.4	1.54	37,191	24.8
	0.25	1.35	31.7	1.55	37,418	24.6
	0.50	1.33	33.0	1.57	37,966	24.0
	0.75	1.31	34.7	1.61	38,823	22.9
Split by n						
	2	1.40	28.4	1.54	36,834	24.9
	3	1.42	27.3	1.55	37,354	24.6
	4	1.43	26.6	1.54	36,910	24.8
	5	1.44	25.9	1.55	37,365	24.6
Vary wealth by factor						
	0.50	1.51	23.0	1.67	38,953	21.2
	0.75	1.43	26.9	1.60	37,844	23.1
	0.90	1.40	28.6	1.56	37,366	24.1
	1.00	1.38	29.9	1.54	37,053	24.8
	1.10	1.36	31.1	1.52	36,769	25.5
	1.25	1.34	33.0	1.49	36,431	26.5
	1.50	1.30	36.1	1.46	35,908	28.1
Fix wmin at level						
	200,000	1.36	35.4	1.42	130,263	24.8
	300,000	0.91	147.4	1.30	121,892	22.7
	500,000	0.77	317.6	1.25	159,465	NaN
	750,000	0.75	349.5	1.22	180,695	NaN
	1,000,000	0.74	368.0	1.22	518,935	NaN
	1,500,000	0.72	434.7	1.60	10,454,508	NaN
	2,000,000	0.71	451.5	1.67	12,399,060	NaN
Fix wmin at percentile						
	0.40	1.38	29.8	1.54	37,743	24.9
	0.50	1.39	29.0	1.51	40,410	25.5
	0.75	1.40	29.5	1.42	56,490	27.2
	0.90	1.37	33.2	1.51	128,531	26.6
	0.99	0.74	369.4	1.23	565,470	NaN

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.11: Sensitivity Analysis: LV

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.41	30.6	1.66	27,182	23.2
Drop n highest						
	1	1.42	30.3	1.67	27,332	22.9
	2	1.42	30.2	1.68	27,403	22.7
	5	1.43	29.2	1.70	27,702	22.2
	10	1.45	28.5	1.73	28,061	21.7
Drop top fraction						
	0.01	1.42	30.3	1.67	27,332	22.9
	0.05	1.43	29.6	1.69	27,593	22.4
	0.10	1.45	28.3	1.72	27,912	21.9
	0.25	1.44	28.9	1.74	28,116	21.3
	0.50	1.45	28.5	1.82	29,154	20.2
Drop n lowest						
	1	1.41	30.6	1.66	27,184	23.2
	2	1.41	30.5	1.66	27,184	23.1
	5	1.41	30.5	1.66	27,198	23.1
	10	1.41	30.4	1.66	27,213	23.0
Drop bottom fraction						
	0.10	1.41	30.6	1.66	27,206	23.1
	0.25	1.42	30.0	1.66	27,221	23.0
	0.50	1.42	30.0	1.68	27,362	22.5
	0.75	1.41	30.4	1.71	27,585	21.8
Split by n						
	2	1.45	28.4	1.66	27,307	23.1
	3	1.48	26.8	1.66	27,350	23.1
	4	1.51	25.4	1.66	27,324	23.1
	5	1.51	25.6	1.67	27,536	23.0
Vary wealth by factor						
	0.50	1.50	25.7	1.74	27,897	21.5
	0.75	1.45	28.2	1.69	27,525	22.3
	0.90	1.43	29.4	1.67	27,304	22.8
	1.00	1.41	30.6	1.66	27,182	23.2
	1.10	1.40	31.4	1.64	27,084	23.5
	1.25	1.39	32.3	1.62	26,905	24.0
	1.50	1.35	34.7	1.59	26,600	24.8
Fix wmin at level						
	200,000	1.40	34.0	1.73	168,531	24.7
	300,000	1.41	34.3	1.73	225,319	23.9
	500,000	1.46	32.3	1.72	262,579	21.0
	750,000	1.44	34.4	1.79	944,641	46.8
	1,000,000	1.54	29.4	2.24	3,590,988	47.8
	1,500,000	1.56	29.0	2.57	5,667,018	NaN
	2,000,000	1.56	29.2	2.57	5,861,792	NaN
Fix wmin at percentile						
	0.40	1.39	32.9	1.90	26,364	20.7
	0.50	1.40	31.5	1.76	26,205	22.1
	0.75	1.43	29.3	1.73	41,020	22.7
	0.90	1.45	29.5	1.68	64,081	22.7
	0.99	1.43	33.8	1.73	278,833	23.2

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.12: Sensitivity Analysis: NL

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.33	38.5	1.5	173,443	28.5
Drop n highest						
	1	1.33	38.4	1.52	175,035	28.0
	2	1.33	38.4	1.53	177,691	27.6
	5	1.33	38.2	1.56	180,798	27.1
	10	1.34	37.7	1.58	183,883	26.7
Drop top fraction						
	0.01	1.33	38.0	1.56	181,339	27.1
	0.05	1.36	36.1	1.63	190,270	25.5
	0.10	1.39	33.9	1.71	199,693	24.1
	0.25	1.45	30.4	2.28	272,642	19.1
	0.50	1.56	25.0	3.86	360,967	14.9
Drop n lowest						
	1	1.33	38.5	1.5	173,476	28.5
	2	1.33	38.5	1.5	173,542	28.5
	5	1.33	38.5	1.5	173,590	28.5
	10	1.33	38.5	1.5	173,743	28.4
Drop bottom fraction						
	0.10	1.32	38.7	1.51	175,487	28.3
	0.25	1.32	38.8	1.53	179,344	27.9
	0.50	1.32	39.4	1.56	184,757	27.1
	0.75	1.30	40.7	1.62	189,922	25.7
Split by n						
	2	1.36	35.6	1.47	160,550	29.0
	3	1.39	33.7	1.43	147,243	29.6
	4	1.41	32.5	1.41	140,114	29.8
	5	1.42	31.7	1.41	138,374	29.8
Vary wealth by factor						
	0.50	1.49	28.0	1.66	185,343	24.5
	0.75	1.39	33.7	1.57	178,371	26.5
	0.90	1.35	36.7	1.53	175,375	27.7
	1.00	1.33	38.5	1.50	173,443	28.5
	1.10	1.31	40.3	1.48	171,999	29.2
	1.25	1.28	42.8	1.46	171,137	30.3
	1.50	1.24	46.6	1.42	170,936	31.9
Fix wmin at level						
	200,000	1.33	38.7	1.59	167,638	27.5
	300,000	1.33	38.6	1.45	179,598	29.4
	500,000	1.28	43.0	1.34	238,018	30.0
	750,000	1.25	47.2	1.34	424,052	31.0
	1,000,000	1.23	50.5	1.33	583,957	30.9
	1,500,000	1.14	63.3	1.32	884,729	31.0
	2,000,000	1.08	77.5	1.33	1,314,889	31.4
Fix wmin at percentile						
	0.40	1.29	55.2	2.11	169,157	21.1
	0.50	1.31	44.2	1.93	163,530	23.1
	0.75	1.33	38.5	1.56	168,842	27.8
	0.90	1.29	41.6	1.36	213,995	29.8
	0.99	1.09	75.3	1.33	1,207,091	30.3

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.13: Sensitivity Analysis: PL

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.63	19.6	2.08	63,121	13.5
Drop n highest						
	1	1.64	19.5	2.10	63,276	13.4
	2	1.64	19.3	2.12	63,377	13.3
	5	1.66	18.7	2.14	63,527	13.1
	10	1.68	18.2	2.16	63,660	13.0
Drop top fraction						
	0.01	1.64	19.5	2.10	63,276	13.4
	0.05	1.66	18.7	2.14	63,527	13.1
	0.10	1.68	18.2	2.16	63,660	13.0
	0.25	1.74	16.9	2.19	63,861	12.8
	0.50	1.82	15.1	2.22	64,056	12.6
Drop n lowest						
	1	1.63	19.6	2.08	63,124	13.5
	2	1.63	19.6	2.08	63,128	13.5
	5	1.63	19.7	2.09	63,143	13.5
	10	1.63	19.7	2.09	63,164	13.5
Drop bottom fraction						
	0.10	1.63	19.7	2.09	63,164	13.5
	0.25	1.63	19.8	2.10	63,278	13.4
	0.50	1.73	17.1	2.11	63,460	13.3
	0.75	1.81	15.4	2.14	63,712	13.1
Split by n						
	2	1.64	19.4	2.08	62,826	13.6
	3	1.68	18.3	2.07	62,594	13.6
	4	1.71	17.5	2.07	62,403	13.6
	5	1.74	16.9	2.06	62,269	13.6
Vary wealth by factor						
	0.50	1.77	16.1	2.16	63,690	12.9
	0.75	1.70	17.7	2.12	63,394	13.2
	0.90	1.66	18.9	2.10	63,227	13.4
	1.00	1.63	19.6	2.08	63,121	13.5
	1.10	1.62	20.1	2.07	63,012	13.6
	1.25	1.60	20.8	2.05	62,835	13.8
	1.50	1.56	22.0	2.02	62,569	14.1
Fix wmin at level						
	200,000	1.52	24.8	1.57	82,485	15.9
	300,000	1.49	27.7	1.64	160,767	15.9
	500,000	1.47	30.2	1.72	299,833	17.1
	750,000	1.44	32.5	1.52	333,265	NaN
	1,000,000	1.42	34.3	1.44	343,503	NaN
	1,500,000	1.35	39.7	1.43	528,371	NaN
	2,000,000	1.26	48.0	1.43	1,026,029	NaN
Fix wmin at percentile						
	0.40	1.53	23.5	3.42	73,592	9.99
	0.50	1.56	21.7	2.85	69,875	11.1
	0.75	1.63	19.9	1.98	67,873	14.0
	0.90	1.53	24.6	1.58	79,837	15.9
	0.99	1.46	31.1	1.57	307,650	NaN

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.14: Sensitivity Analysis: PT

	Scenario	**Pareto**		**GPareto**		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.47	27.1	1.64	133,952	23.8
Drop n highest						
	1	1.47	26.9	1.65	134,341	23.4
	2	1.48	26.5	1.66	134,593	23.1
	5	1.50	25.6	1.68	134,915	22.8
	10	1.51	25.1	1.69	135,232	22.5
Drop top fraction						
	0.01	1.47	26.9	1.65	134,341	23.4
	0.05	1.48	26.5	1.66	134,593	23.1
	0.10	1.49	25.8	1.67	134,822	22.9
	0.25	1.51	25.1	1.69	135,232	22.5
	0.50	1.53	24.0	1.71	135,541	22.1
Drop n lowest						
	1	1.46	27.2	1.64	133,980	23.7
	2	1.46	27.2	1.64	133,997	23.7
	5	1.46	27.2	1.64	134,079	23.7
	10	1.47	27.0	1.64	134,204	23.6
Drop bottom fraction						
	0.10	1.46	27.2	1.64	134,059	23.7
	0.25	1.47	27.0	1.64	134,204	23.6
	0.50	1.48	26.6	1.66	134,589	23.3
	0.75	1.49	26.0	1.67	134,991	22.9
Split by n						
	2	1.48	26.6	1.63	133,621	23.8
	3	1.49	25.7	1.63	133,248	23.8
	4	1.51	24.8	1.63	132,912	23.8
	5	1.53	24.1	1.63	132,625	23.8
Vary wealth by factor						
	0.50	1.54	23.7	1.68	135,019	22.7
	0.75	1.49	25.9	1.66	134,470	23.2
	0.90	1.47	26.9	1.65	134,170	23.6
	1.00	1.47	27.1	1.64	133,952	23.8
	1.10	1.46	27.3	1.63	133,746	24.0
	1.25	1.46	27.5	1.62	133,433	24.3
	1.50	1.45	27.7	1.60	132,978	24.8
Fix wmin at level						
	200,000	1.52	25.0	1.57	145,440	24.6
	300,000	1.55	24.4	1.48	180,933	25.6
	500,000	1.42	32.0	1.40	261,981	25.8
	750,000	1.40	34.7	1.43	476,667	25.5
	1,000,000	1.37	36.8	1.38	589,345	26.0
	1,500,000	1.36	38.3	1.66	1,519,466	24.4
	2,000,000	1.37	38.0	1.75	2,094,883	35.3
Fix wmin at percentile						
	0.40	1.30	38.7	1.75	88,791	22.4
	0.50	1.36	33.6	1.72	96,440	22.8
	0.75	1.47	27.1	1.64	135,921	23.7
	0.90	1.54	25.2	1.48	196,636	25.7
	0.99	1.36	38.5	1.58	1,209,528	28.5

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

Table E.15: Sensitivity Analysis: SI

	Scenario	Pareto		GPareto		
	Parameter	Alpha	Share top 1%	Shape	Scale	Share top 1%
Baseline						
	NA	1.57	21.9	1.83	100,640	17
Drop n highest						
	1	1.57	21.6	1.85	101,132	16.7
	2	1.58	21.3	1.87	101,530	16.5
	5	1.61	20.5	1.89	102,102	16.2
	10	1.64	19.5	1.92	102,622	15.9
Drop top fraction						
	0.01	1.57	21.6	1.85	101,132	16.7
	0.05	1.61	20.5	1.89	102,102	16.2
	0.10	1.64	19.5	1.92	102,622	15.9
	0.25	1.68	18.3	2.01	106,475	15.1
	0.50	1.75	16.8	2.64	137,790	12.9
Drop n lowest						
	1	1.57	21.9	1.83	100,686	17.0
	2	1.57	21.9	1.83	100,695	17.0
	5	1.57	21.8	1.83	100,810	17.0
	10	1.57	21.8	1.84	100,971	16.9
Drop bottom fraction						
	0.10	1.57	21.8	1.84	100,971	16.9
	0.25	1.58	21.6	1.85	101,427	16.7
	0.50	1.58	21.3	1.89	102,208	16.3
	0.75	1.59	21.0	1.93	103,099	15.8
Split by n						
	2	1.65	19.3	1.81	98,792	17.2
	3	1.68	18.3	1.82	100,202	17.0
	4	1.56	22.2	1.85	102,141	16.9
	5	1.61	20.5	1.85	102,481	16.8
Vary wealth by factor						
	0.50	1.79	15.8	1.94	102,119	15.7
	0.75	1.65	19.1	1.88	101,361	16.4
	0.90	1.60	20.9	1.85	100,933	16.8
	1.00	1.57	21.9	1.83	100,640	17.0
	1.10	1.54	23.0	1.81	100,361	17.3
	1.25	1.51	24.3	1.78	99,965	17.7
	1.50	1.47	26.2	1.74	99,391	18.4
Fix wmin at level						
	200,000	1.56	22.7	1.63	105,682	18.6
	300,000	1.53	24.9	1.54	153,086	19.4
	500,000	1.51	27.3	1.62	308,205	19.3
	750,000	1.49	29.1	1.66	510,503	15.3
	1,000,000	1.49	30.2	1.66	711,638	23.7
	1,500,000	1.50	30.3	1.67	1,233,751	NaN
	2,000,000	1.51	30.0	1.70	2,166,158	NaN
Fix wmin at percentile						
	0.40	1.53	24.2	2.12	83,864	15.1
	0.50	1.55	22.9	1.99	86,051	15.9
	0.75	1.57	21.9	1.73	99,970	17.8
	0.90	1.53	24.7	1.54	145,406	19.2
	0.99	1.48	30.7	1.67	799,155	NaN

Note: This table is based on all five implicates of HFCS 2017 data. NaN reported in case the location parameter of the scenario exceeds the replacement threshold.

E.2 Sensitivity Analysis: w_0

Table E.16 - Sensitivity Analysis w_0 by Country **123**
Table E.29

Table E.16: Sensitivity Analysis w_0: AT

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		67.7016	61.9124	57.3362	50.2014	38.9815	30.6780
Fix w0 at level							
	1,000,000	67.5115	61.9371	57.1748	50.2220	38.8712	30.6911
	1,500,000	67.5129	61.9260	57.1760	50.2126	38.8720	30.6850
	2,000,000	67.5035	61.9045	57.1681	50.1946	38.8666	30.6736
	2,500,000	67.5124	61.9093	57.1756	50.1987	38.8717	30.6762
	5,000,000	67.5465	61.9375	57.2045	50.2223	38.8913	30.6911
Fix w0 at percentile							
	0.80	67.7299	62.2272	57.3600	50.4657	38.9973	30.8460
	0.90	67.5124	61.9583	57.1757	50.2397	38.8718	30.7023
	0.95	67.5280	61.9701	57.1889	50.2496	38.8808	30.7084
	0.99	67.5023	61.9062	57.1670	50.1960	38.8659	30.6745

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.17: Sensitivity Analysis w_0: BE

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		50.8548	51.4463	39.1146	39.5678	21.2647	21.2016
Fix w0 at level							
	1,000,000	49.4141	51.3881	38.0065	39.5205	20.6621	21.1747
	1,500,000	49.4142	51.4562	38.0066	39.5757	20.6622	21.2061
	2,000,000	49.3964	51.4631	37.9928	39.5814	20.6547	21.2093
	2,500,000	49.3812	51.4597	37.9812	39.5786	20.6484	21.2077
	5,000,000	49.3818	51.4833	37.9816	39.5978	20.6486	21.2186
Fix w0 at percentile							
	0.80	49.9092	51.7578	38.3873	39.8214	20.8692	21.3462
	0.90	49.5735	51.5140	38.1291	39.6227	20.7288	21.2328
	0.95	49.3954	51.3838	37.9921	39.5170	20.6543	21.1727
	0.99	49.3734	51.4663	37.9752	39.5840	20.6451	21.2108

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.18: Sensitivity Analysis w_0: DE

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		63.5060	58.5243	52.1019	45.2381	32.9049	24.5981
Fix w0 at level							
	1,000,000	65.0388	58.6038	53.3594	45.3044	33.6989	24.6373
	1,500,000	64.9981	58.5170	53.3260	45.2319	33.6778	24.5944
	2,000,000	65.0279	58.5475	53.3504	45.2574	33.6932	24.6095
	2,500,000	65.0416	58.5596	53.3617	45.2676	33.7003	24.6156
	5,000,000	65.0372	58.5256	53.3581	45.2392	33.6980	24.5988
Fix w0 at percentile							
	0.80	65.2807	59.1976	53.5742	45.8711	33.8586	25.0595
	0.90	65.1199	58.7730	53.4260	45.4476	33.7409	24.7230
	0.95	65.0298	58.6053	53.3520	45.3055	33.6942	24.6379
	0.99	65.0391	58.5622	53.3596	45.2698	33.6990	24.6169

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.19: Sensitivity Analysis w_0: FI

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		45.6237	47.7019	32.9429	34.0929	15.4658	15.2431
Fix w0 at level							
	1,000,000	47.3653	47.6932	34.2003	34.0844	16.0561	15.2381
	1,500,000	47.3609	47.6996	34.1972	34.0907	16.0547	15.2418
	2,000,000	47.3579	47.6990	34.1950	34.0901	16.0536	15.2415
	2,500,000	47.3603	47.7047	34.1968	34.0957	16.0545	15.2448
	5,000,000	47.3603	47.7059	34.1967	34.0969	16.0544	15.2456
Fix w0 at percentile							
	0.80						
	0.90	47.4036	47.7454	34.2280	34.1370	16.0691	15.2704
	0.95	47.3821	47.7027	34.2125	34.0937	16.0618	15.2436
	0.99	47.3594	47.6986	34.1961	34.0897	16.0542	15.2412

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.20: Sensitivity Analysis w_0: FR

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		49.4492	52.8102	36.9068	40.4210	18.7111	22.0354
Fix w0 at level							
	1,000,000	50.9395	52.8290	38.0191	40.4383	19.2750	22.0467
	1,500,000	50.9467	52.8449	38.0245	40.4531	19.2777	22.0564
	2,000,000	50.9424	52.8369	38.0213	40.4457	19.2761	22.0515
	2,500,000	50.9365	52.8255	38.0169	40.4351	19.2739	22.0446
	5,000,000	50.9324	52.8151	38.0139	40.4255	19.2724	22.0383
Fix w0 at percentile							
	0.80						
	0.90	50.8331	52.6195	37.9397	40.2513	19.2348	21.9274
	0.95	50.9286	52.8031	38.0110	40.4146	19.2709	22.0312
	0.99	50.9469	52.8449	38.0246	40.4532	19.2778	22.0565

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.21: Sensitivity Analysis w_0: HU

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		53.5026	52.9025	42.3602	41.0199	24.6330	22.2924
Fix w0 at level							
	1,000,000	53.4674	52.9017	42.3324	41.0188	24.6169	22.2915
	1,500,000	53.4645	52.9036	42.3301	41.0218	24.6156	22.2942
	2,000,000	53.4663	52.9028	42.3315	41.0204	24.6164	22.2929
	2,500,000	53.4672	52.9024	42.3323	41.0197	24.6168	22.2922
	5,000,000	53.4675	52.9023	42.3325	41.0194	24.6169	22.2920
Fix w0 at percentile							
	0.80	53.2143	52.8188	42.1326	40.9706	24.5016	22.2753
	0.90	53.4076	52.8923	42.2853	41.0163	24.5897	22.2932
	0.95	53.4643	52.8895	42.3299	41.0038	24.6155	22.2800
	0.99	53.4676	52.9010	42.3326	41.0178	24.6170	22.2908

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.22: Sensitivity Analysis w_0: IE

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		59.0412	56.2899	47.2573	43.7469	28.1820	24.2788
Fix w0 at level							
	1,000,000	50.6590	54.4122	46.8349	42.0457	27.2675	23.2780
	1,500,000	51.2261	56.0458	47.1945	43.5574	28.1445	24.1702
	2,000,000	51.2346	56.3076	47.3608	43.7611	28.2436	24.2869
	2,500,000	51.2538	56.2776	47.3022	43.7381	28.2086	24.2739
	5,000,000	51.2778	56.2731	47.2285	43.7354	28.1647	24.2732
Fix w0 at percentile							
	0.80	58.4069					
	0.90	58.2762	53.0387	46.5331	40.6705	26.2532	22.2918
	0.95	51.3160	55.8259	47.0807	43.3842	28.0765	24.0691
	0.99	51.2671	56.2861	47.3526	43.7453	28.2386	24.2784

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.23: Sensitivity Analysis w_0: IT

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		49.1542	44.7024	37.4667	31.5018	19.9460	13.2782
Fix w0 at level							
	1,000,000	49.3484	44.6966	37.6147	31.4958	20.0248	13.2747
	1,500,000	49.3482	44.7096	37.6145	31.5092	20.0247	13.2826
	2,000,000	49.3479	44.7103	37.6144	31.5098	20.0246	13.2830
	2,500,000	49.3484	44.7047	37.6147	31.5041	20.0248	13.2796
	5,000,000	49.3464	44.7068	37.6132	31.5063	20.0240	13.2809
Fix w0 at percentile							
	0.80	49.0932	44.7983	37.4202	31.6039	19.9212	13.3400
	0.90	49.3114	44.7149	37.5865	31.5146	20.0098	13.2859
	0.95	49.3426	44.6963	37.6103	31.4955	20.0224	13.2745
	0.99	49.3483	44.7112	37.6147	31.5108	20.0247	13.2836

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.24: Sensitivity Analysis w_0: LT

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		56.1707	53.9946	46.4578	42.9707	29.8962	24.8472
Fix w0 at level							
	1,000,000	56.0593	54.0086	46.3656	42.9826	29.8368	24.8547
	1,500,000	56.0484	53.9910	46.3566	42.9676	29.8310	24.8453
	2,000,000	56.0456	53.9867	46.3543	42.9639	29.8295	24.8430
	2,500,000	56.0472	53.9900	46.3556	42.9667	29.8303	24.8448
	5,000,000	56.0496	53.9947	46.3576	42.9708	29.8316	24.8473
Fix w0 at percentile							
	0.80	56.2112	54.0452	46.4913	43.0139	29.9176	24.8741
	0.90	56.0945	53.9430	46.3948	42.9265	29.8555	24.8198
	0.95	56.0916	54.0296	46.3923	43.0006	29.8540	24.8658
	0.99	56.0582	54.0091	46.3647	42.9831	29.8362	24.8550

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.25: Sensitivity Analysis w_0: LV

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		59.8243	55.0890	48.8758	42.7551	30.5681	23.1697
Fix w0 at level							
	1,000,000	58.7473	55.0788	47.9959	42.7463	30.0178	23.1645
	1,500,000	58.7514	55.0839	47.9992	42.7507	30.0199	23.1670
	2,000,000	58.7555	55.0894	48.0026	42.7553	30.0220	23.1698
	2,500,000	58.7578	55.0922	48.0045	42.7577	30.0232	23.1712
	5,000,000	58.7595	55.0933	48.0059	42.7586	30.0240	23.1718
Fix w0 at percentile							
	0.80	59.0583	55.0561	48.2500	42.7271	30.1767	23.1530
	0.90	58.9600	55.2540	48.1697	42.8947	30.1264	23.2525
	0.95	58.8581	55.1887	48.0864	42.8394	30.0744	23.2197
	0.99	58.7527	55.0870	48.0003	42.7533	30.0205	23.1686

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.26: Sensitivity Analysis w_0: NL

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		67.8805	61.4719	57.2291	48.7565	38.5035	28.4713
Fix w0 at level							
	1,000,000	69.2750	61.4721	58.4050	48.7559	39.2948	28.4704
	1,500,000	69.3138	61.5015	58.4377	48.7811	39.3169	28.4862
	2,000,000	69.3251	61.5036	58.4473	48.7837	39.3233	28.4882
	2,500,000	69.3341	61.5080	58.4548	48.7871	39.3284	28.4901
	5,000,000	69.3576	61.5243	58.4746	48.8007	39.3417	28.4983
Fix w0 at percentile							
	0.80	69.5047	62.3324	58.5906	49.4154	39.4072	28.6348
	0.90	69.3366	61.6122	58.4568	48.8495	39.3296	28.5114
	0.95	69.2628	61.4914	58.3947	48.7673	39.2879	28.4745
	0.99	69.3249	61.5090	58.4470	48.7876	39.3232	28.4903

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.27: Sensitivity Analysis w_0: PL

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		47.8657	42.3417	36.5780	30.2460	19.5889	13.5215
Fix w0 at level							
	1,000,000	47.8053	42.3552	36.5319	30.2572	19.5642	13.5274
	1,500,000	47.8088	42.3581	36.5345	30.2596	19.5656	13.5286
	2,000,000	47.8091	42.3555	36.5348	30.2575	19.5657	13.5275
	2,500,000	47.8093	42.3539	36.5349	30.2561	19.5658	13.5268
	5,000,000	47.8100	42.3518	36.5354	30.2544	19.5661	13.5259
Fix w0 at percentile							
	0.80	47.8575	42.4841	36.5717	30.3657	19.5855	13.5854
	0.90	47.8014	42.3417	36.5288	30.2460	19.5626	13.5215
	0.95	47.7732	42.2895	36.5073	30.2030	19.5510	13.4990
	0.99	47.7999	42.3501	36.5277	30.2530	19.5620	13.5252

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.28: Sensitivity Analysis w_0: PT

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		56.3187	55.5002	45.2017	43.3167	27.1312	23.7597
Fix w0 at level							
	1,000,000	56.4067	55.4927	45.2723	43.3074	27.1736	23.7525
	1,500,000	56.4086	55.4923	45.2739	43.3070	27.1745	23.7522
	2,000,000	56.4058	55.5011	45.2716	43.3178	27.1732	23.7606
	2,500,000	56.4062	55.5020	45.2719	43.3187	27.1734	23.7612
	5,000,000	56.4043	55.5071	45.2704	43.3251	27.1725	23.7662
Fix w0 at percentile							
	0.80	56.0090	55.4373	44.9546	43.3059	26.9849	23.7769
	0.90	56.3489	55.4935	45.2261	43.3137	27.1462	23.7595
	0.95	56.3901	55.4900	45.2591	43.3065	27.1658	23.7528
	0.99	56.4094	55.4920	45.2745	43.3065	27.1749	23.7518

Note: This table is based on all five implicates of HFCS 2017 data.

Table E.29: Sensitivity Analysis w_0: SI

	Parameter	Top 10% Share		Top 5% Share		Top 1% Share	
		Pareto	G-Pareto	Pareto	G-Pareto	Pareto	G-Pareto
Baseline							
		50.4033	46.4612	39.2314	34.5193	21.9257	17.0311
Fix w0 at level							
	1,000,000	50.3375	46.4702	39.1803	34.5277	21.8971	17.0361
	1,500,000	50.3363	46.4670	39.1793	34.5248	21.8965	17.0343
	2,000,000	50.3372	46.4719	39.1800	34.5291	21.8969	17.0369
	2,500,000	50.3377	46.4687	39.1804	34.5260	21.8972	17.0350
	5,000,000	50.3381	46.4721	39.1808	34.5291	21.8974	17.0368
Fix w0 at percentile							
	0.80	50.2604	46.4500	39.1203	34.5246	21.8636	17.0403
	0.90	50.3379	46.4311	39.1806	34.4935	21.8973	17.0162
	0.95	50.3297	46.4046	39.1742	34.4691	21.8937	17.0017
	0.99	50.3365	46.4662	39.1794	34.5241	21.8966	17.0340

Note: This table is based on all five implicates of HFCS 2017 data.